ESTIMATING ORAL ANTICOAGULANT COMPARATIVE EFFECTIVENESS IN THE SETTING OF EFFECT HETEROGENEITY: COMPARING CLINICAL TRIAL TRANSPORT AND NON-EXPERIMENTAL EPIDEMIOLOGIC METHODS

Michael Webster-Clark

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Epidemiology in the Gillings School of Global Public Health.

Chapel Hill
2019

Approved by:

Jennifer Lund

Til Stürmer

Charles Poole

Jess Edwards

Ross Simpson

ABSTRACT

Michael Webster-Clark: Estimating Oral Anticoagulant Comparative Effectiveness in the Setting of Effect Heterogeneity: Comparing Clinical Trial Transport and Non-experimental Epidemiologic Methods
(Under the direction of Jennifer Lund)

Oral anticoagulation is vital to the health of patients with atrial fibrillation at elevated risk of stroke. The first treatment for these patients, warfarin, was approved in the 1990s. Since 2010, dabigatran has been available for use after demonstrating non-inferiority to warfarin in a randomized controlled trial. Non-experimental studies comparing dabigatran to warfarin and censoring at treatment discontinuation have shown greater benefits than the original trial for all-cause mortality and attenuated harms for gastrointestinal bleeding.

The goals of this dissertation, then, were to compute and compare 1) estimates of the absolute-scale effects of dabigatran vs warfarin initiation on ischemic stroke (IS), death, and gastrointestinal bleeding (GIB) in trial-eligible older adults using non-experimental Medicare data and 2) estimates of those effects in the same populations using inverse odds of sampling weights to transport results from the Randomized Evaluation of Long-Term Anticoagulation (RE-LY) trial.

First, we conducted a propensity score weighted non-experimental study with the new user active comparator design in a 20% random sample of Medicare beneficiares. We estimated on-treatment two-year risk differences for IS (RD for dabigatran users, RD_{dabi}: -0.67%, 95% CI -1.10%, -0.24%), mortality (RD_{dabi}: -2.98%, 95% CI -3.97%, -1.95%) and GIB (RD_{dabi}: 0.51%, 95% CI -0.30%, 1.31%). Intention-to-treat estimates showed attenuation for mortality (RD_{dabi}: -1.65%, 95% CI -2.32%, -0.98%) and reversal for IS (RD_{dabi}: 0.16%, 95% CI -0.20%, 0.52%).

Next, we reweighted RE-LY to resemble the Medicare new users of warfarin or dabigatran (restricted to those with less than 15% predicted probability of frailty). After weighting, we estimated

on-treatment two-year risk differences for IS (RD_{dabi}: -0.77%, 95% CI -1.69%, 0.14%), death (RD_{dabi}: -0.57%, 95% CI -1.83%, 0.68%) and GIB (RD_{dabi}: 1.75%, 95% CI 0.76%, 2.74%).

These twin studies show non-experimental and weighted trial analyses comparing dabigatran to warfarin agree much better for IS than they do for mortality or GIB. This could be due to confounding in the non-experimental estimates, missing treatment effect modifiers, or outcome misclassification. Researchers should be cautious about comparing studies without considering treatment effect heterogeneity and differences in adherence across study populations.

To my family and all my friends: thank you for all your help and support. I could never have done this without you.

ACKNOWLEDGEMENTS

I would like to acknowledge the efforts of everyone on my dissertation committee for supporting me throughout the planning, analysis, and synthesis stages of this project. I would also like to thank the Royster Society of Fellows for providing me critical financial support throughout my time at the University of North Carolina at Chapel Hill, and Virginia Pate for providing critical assistance in the cohort creation and analysis stages of the project. Finally, I would like to acknowledge the help and assistance of Boehringer Ingelheim and Clinical Study Data Request™, as well as the Clinical Trial Data Transparency Platform, for their efforts to make individual-level data available for research projects like this one.

TABLE OF CONTENTS

LIST OF TABLES xi

LIST OF FIGURES xiii

LIST OF ABBREVIATIONS xv

CHAPTER 1: SPECIFIC AIMS 1

CHAPTER 2: BACKGROUND AND SIGNIFICANCE 3

2.1: Atrial Fibrillation 3

2.1.1: Pathophysiology 3

2.1.2: Population burden 5

2.2: Pharmacologic Treatment Options 7

2.2.1: Warfarin 7

2.2.2: Novel Oral Anticoagulants 8

2.2.2.1: Dabigatran 9

2.2.2.2: Other NOACs 10

2.3: Existing Evidence on Dabigatran Safety and Efficacy 11

2.3.1: Randomized Controlled Trials 11

2.3.1.1: RE-LY Trial 11

2.3.1.2: Heterogeneity in the RE-LY trial 12

2.3.1.3: Other Trials of Dabigatran vs. Warfarin 13

2.3.2: Non-experimental Studies 15

2.4: Heterogeneity Between Trials and Clinical Cohorts 23

2.4.1: Demographic Characteristic Heterogeneity 23
2.4.2: Warfarin Management Heterogeneity 24
2.4.3: Medication Adherence and Persistence Heterogeneity 25
2.5: Methods for Transporting Causal Effects 27
2.5.1: Standardization Methods 27
2.5.2: Weighting Methods 28
2.6: Methods for Assessing Transportability 30
2.6.1: Comparing Marginal Variable Distributions 30
2.6.2: Assessing Proportion of Patients Eligible for Trials 31
2.6.3: Weighting Methods for Assessing Transportability 32
2.7: Public Health Significance 35
CHAPTER 3: RESEARCH PLAN AND METHODS 36
3.1: Overview 36
3.2: Data Sources and Collection 37
3.2.1: RE-LY Population 37
3.2.1.1: RE-LY Population Description 37
3.2.1.2: Outcome Assessment in RE-LY 38
3.2.1.3: Exposure Assessment in RE-LY 39
3.2.1.4: Covariate Assessment in RE-LY 40
3.2.2: Medicare Atrial Fibrillation Population 43
3.2.2.1: Medicare Atrial Fibrillation Population Description 43
3.2.2.2: Outcome Assessment in Medicare 45
3.2.2.3: Exposure Assessment in Medicare 46
3.2.2.4: Covariate Assessment in Medicare 47

3.2.3: Data Combination 49

3.3: Data Analysis 51

3.3.1: Aim 1, Estimating Non-experimental Treatment Effects in Medicare 51

3.3.1.1: Aim 1.1, New User Active Comparator Design 51

3.3.2: Aim 2, Transporting Treatment Effects with Inverse Odds of Sampling Weights 53

3.3.2.1: Aim 2.1, Assessing Transportability 53

3.3.2.2: Aim 2.2, Transporting Treatment Effects 56

3.3.2.3: Aim 2.3, Contrasting Effect Estimates 57

3.4. Human Subjects 59

CHAPTER 4: STAYING ON TREATMENT MATTERS: ESTIMATING EFFECTS OF DABIGATRAN VERSUS WARFARIN IN MEDICARE 60

4.1. Introduction 60

4.2. Methods 62

4.2.1. Study Population 62

4.2.2. Exposure 62

4.2.3. Outcomes 63

4.2.4. Covariates 64

4.2.5. Statistical Analyses 64

4.2.6. Sensitivity Analyses 65

4.3. Results 66

4.4. Discussion 69

CHAPTER 5: REWEIGHTING ORANGES TO APPLES: COMPARING TRANSPORTED RE-LY TRIAL AND NON-EXPERIMENTAL ESTIMATES IN ATRIAL FIBRILLATION 85

5.1. Introduction 85

5.2. Methods 88

5.2.1. Parameters of Interest 88
5.2.2. Study Populations 88
5.2.3. Exposure 90
5.2.4. Outcomes 91
5.2.5. Covariates 91
5.2.6. Statistical Analyses 93
5.2.7. Sensitivity Analyses 95
5.3. Results 96
5.4. Discussion 100
CHAPTER 6: CONCLUSIONS 122
6.1. Main Findings 122
6.2. Significance 125
6.3. Future Directions 126
APPENDIX A: OUTCOME CODING 128
APPENDIX B: ELIGIBILITY CRITERIA AND COVARIATES 132
APPENDIX C: PROCEDURE CODES FOR ANTICOAGULATION MANAGEMENT 136
REFERENCES 137

LIST OF TABLES

Table 1: Published studies comparing warfarin vs. dabigatran 22

Table 2: Aim 2 cohort summary 75

Table 3: Person-years, events, and risks in Aim 1 analyses 77

Table 4: Risk ratios and risk differences in Aim 1 main outcome analyses 81

Table 5: Risk ratios and risk differences in Aim 1 secondary outcome analyses 82

Table 6: Two-year risks of outcomes across sensitivity analyses 83

Table 7: Two-year risk differences for outcomes across sensitivity analyses 84

Table 8: Trial and main target population covariate distributions 104

Table 9: Trial and unrestricted target population covariate distributions 104

Table 10: Results from analyses in trial patients over 65 106

Table 11: Sampling model covariate distributions (main target population) 107

Table 12: Sampling model covariate distributions (secondary target population) 107

Table 13: Main outcome person-years and events after weights 114

Table 14: Other outcome person-years and events after weights 115

Table 15: Person-years and events after secondary target population weights 116

Table 16: SMR-weighted non-experimental risk differences in the target populations 117

Table 17: Transportability assessment results for the main target populations 120

Table 18: Transportability assessment results for the secondary target populations 120

Table 19: Risk differences when transporting to sensitivity analysis target populations 121

Table 20: Risk differences with alternate sampling models 121

LIST OF FIGURES

Figure 1: The heart and atrial fibrillation 3

Figure 2: Atrial fibrillation trends over time 5

Figure 3: Warfarin as rat poison 7

Figure 4: Dabigatran packaging 9

Figure 5: Subgroup results from the RE-LY trial 12

Figure 6: Inverse probability of sampling weights visual representation 28

Figure 7: Survival curves as used in transportability assessment 32

Figure 8: Causal diagram examining sampling and ischemic stroke 40

Figure 9: Causal diagram examining sampling and gastrointestinal bleeding 41

Figure 10: Exposure identification and follow-up 46

Figure 11: Causal diagram examining treatment and outcomes 47

Figure 12: Data flow diagram for the combination of each population's data 49

Figure 13: Aim 1 conceptual diagram 51

Figure 14: Aim 2 conceptual diagram 53

Figure 15: Example transportability assessment cumulative incidence curves 53

Figure 16: Aim 1 cohort flow diagram 74

Figure 17: Aim 1 cohort standardized mean differences 76

Figure 18: IPTW ischemic stroke cumulative incidence curves 78

Figure 19: IPTW all-cause mortality cumulative incidence curves 78

Figure 20: IPTW gastrointestinal bleeding cumulative incidence curves 79

Figure 21: IPTW all stroke cumulative incidence curves 79

Figure 22: IPTW major bleeding cumulative incidence curves 80

Figure 23: Aim 2 cohort flow diagram 103

Figure 24: Aim 2 cohort standardized mean differences (trial vs target) 105

Figure 25: Ischemic stroke cumulative incidence curves 108

Figure 26: Ischemic stroke cumulative incidence curves (unrestricted targets) 109

Figure 27: All-cause mortality cumulative incidence curves 110

Figure 28: Gastrointestinal bleeding cumulative incidence curves 111

Figure 29: All stroke cumulative incidence curves 112

Figure 30: Major bleeding cumulative incidence curves 113

Figure 31: Constrasting treatment effects: main target populations 118

Figure 32: Contrasting treatment effects: unrestricted target populations 119

LIST OF ABBREVIATIONS

AF	Atrial fibrillation
$CHADS_2$	Stroke risk score based upon Congestive heart failure, Hypertension, Age, Diabetes, and past stroke
C.I.	Confidence interval
CSDR	Clinical Study Data Request
CTDT	Clinical Trial Data Transparency (platform)
EMM	Effect measure modification or effect measure modifier
HR	Hazard ratio
NOAC	Novel oral anticoagulant
RCT	Randomized controlled trial
RD	Risk difference
RR	Risk ratio
sIOSW	Stabilized inverse odds of sampling weights
sIPCW	Stabilized inverse probability of censoring weights
sIPTW	Stabilized inverse probability of treatment weights

CHAPTER 1: SPECIFIC AIMS

Atrial fibrillation affects 33 million adults worldwide.[1] Even if individuals with atrial fibrillation are asymptomatic, stroke incidence in the atrial fibrillation population is much higher and resulting strokes are more frequently associated with death, hospitalization, and long-term disability than strokes in adults without atrial fibrillation.[2,3] Warfarin, the standard treatment for preventing strokes in atrial fibrillation, is difficult to manage therapeutically due to its lengthy half-life and narrow therapeutic range. Warfarin overdose can also result in catastrophic bleeding events.[4] Novel oral anticoagulants have been shown to be non-inferior to warfarin administered with systematic management protocols in clinical trial populations.[5] One of the first novel anticoagulants to be approved in the United States, dabigatran, was shown to be more effective than warfarin at stroke and embolic event prevention (HR 0.66, 95% C.I. 0.53-0.82) with no increase in bleeding (HR 0.93, 95% C.I. 1.07) in the RE-LY trial.[6]

However, estimates of efficacy in these clinical trials are not estimates of effectiveness in clinical care.[7,8] Patients selected into trials tend to be young with fewer comorbidities than the general population; this can modify the population average treatment effect.[9,10] To address concerns about this potential treatment effect modification, studies have used observational claims data to directly estimate safety of novel oral anticoagulants compared to warfarin in clinical care and observed attenuated efficacy and differing safety profiles. Unfortunately, their results may be confounded by unmeasured variables.[11-20] Furthermore, warfarin management protocols from trials do not necessarily represent the way warfarin is managed for patients in routine clinical care, making it difficult to know how consistent warfarin treatment is between trial and observational populations.[21]

The overarching goal of this dissertation is to estimate treatment effects in various Medicare populations of older adults, juxtaposing the estimates obtained from relatively novel methods for transporting treatment effects from a trial with estimates obtained using more standard propensity score weighting in those older adults.

Specifically, I will:

Specific Aim 1: Estimate the effect of dabigatran versus warfarin initiation on two-year risks of ischemic stroke, death, and gastrointestinal bleeding in Medicare beneficiaries with atrial fibrillation using non-experimental data and propensity score methods.

1.1: Estimate absolute effects of dabigatran versus warfarin on two-year risks of ischemic stroke, death, and gastrointestinal bleeding in warfarin and dabigatran new users with a new user active comparator design under both intention-to-treat and as-treated follow-up.

Specific Aim 2: Estimate the effect of dabigatran versus warfarin initiation on the two-year risk of ischemic stroke, death, and gastrointestinal bleeding among Medicare beneficiaries with atrial fibrillation using transportability methods reweighting the RE-LY trial data.

2.1: Compare risks of each outcome between the RE-LY and Medicare cohorts on each treatment before and after using inverse odds of sampling weights to standardize RE-LY to Medicare with respect to effect measure modifiers (EMM).

2.2: Estimate the absolute effect of dabigatran versus warfarin on two-year risks of ischemic stroke, death, and gastrointestinal bleeding using inverse odds of sampling weights to standardize the RE-LY trial to the Medicare cohort in effect measure modifiers.

2.3: Compare these estimates to those from the the non-experimental work.

CHAPTER 2: BACKGROUND AND SIGNIFICANCE

2.1: Atrial Fibrillation

2.1.1: Pathophysiology

Atrial fibrillation (AF) is a type of sustained disorder of cardiac rhythm in which the upper chambers of the heart beat irregularly due to abnormal impulse formation for any reason (see **Figure 1**).[22] AF is typically divided into valvular and non-valvular categories based upon the etiology of the rhythm disorder, with valvular diseases tracing back to rheumatic diseases of the mitral valve or valve replacement.[23] In both types of AF the atria's chaotic beat overwhelms the atrioventricular node until it is unable to create consistent ventricular contractions, creating irregular time intervals between heartbeats. This leads to the symptoms of AF including heart palpitations, lightheadedness, fatigue, shortness of breath, and chest pain, though these symptoms do not occur in everyone with an irregular heartbeat.[22] Symptomatic episodes may be brought on by a

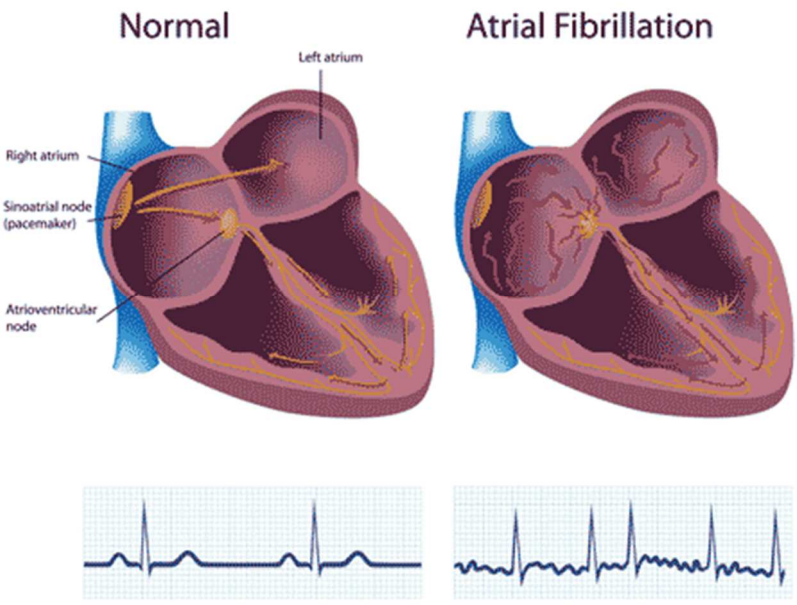

Figure 1: A heart and EKG with and without AF-picture from PracticalClinicalSkills.com. Note the chaotic lines within the heart on the right and the irregular spacing between EKG peaks.

variety of stimuli, including caffeine, stress, poor sleep, and (for some individuals) exercise, but the chronic disease persists regardless. The best ways to relieve symptoms are the use of medications or procedures for rate or rhythm control, but neither of these strategies have substantive morbidity or mortality benefits.[24]

Even when symptoms are controlled, the most devastating consequence of AF is stroke after an atrial blood clot, or thrombus, is dislodged and finds its way to the brain. Rudolf Virchow proposed a trio of risk factors for thrombotic events: abnormal changes in the walls of blood vessels, abnormal blood flow, and abnormal blood constraints.[25] Research using modern imaging and other technology has shown that AF leads to long-term changes in the heart and blood's structure that suffice to cover all three components of Virchow's Triad to substantially promote thrombogenesis.[25] As a result of these structural and chemical changes in the human body, AF increases an individual's risk of stroke by a factor of 5 and creates associated increases in rates of dementia and mortality; the strokes themselves are also more harmful on average than strokes in individuals without AF.[2,26] This is especially concerning because many of the risk factors for AF (high blood pressure, congestive heart failure, history of heart attacks, obesity, and diabetes) are also risk factors for stroke.[27] While guidelines disagree about which AF patients merit oral anticoagulation to counteract this increased stroke risk, most agree that aging individuals with other comorbidities that increase the risk of stroke should at least consider their specific risk-benefit profile.[28]

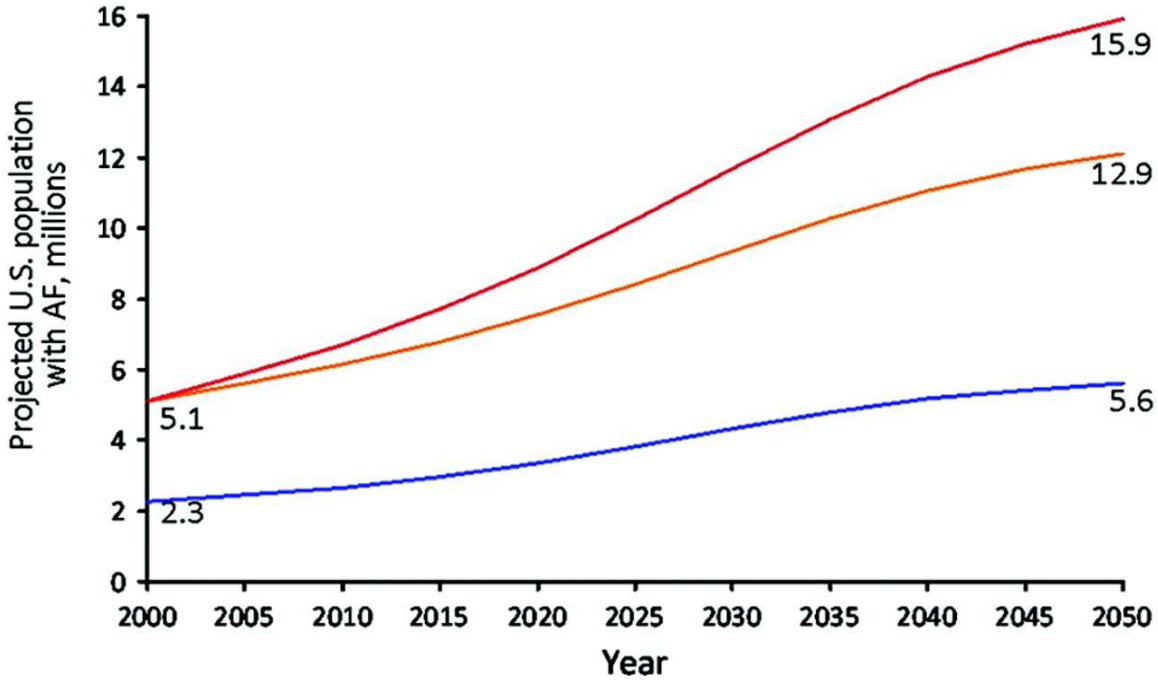

Figure 2: Projected U.S. population with AF over time from Miyasaka et al's work in Minnesota.

2.1.2: Population burden

AF is one of the most common heart rhythm disorders, affecting 33 million adults around the world.[1] In the United States, between 2.7 and 6.1 million people have atrial fibrillation, with it affecting 2% of individuals under 65 and 9% of individuals over 65.[22] As the world's population ages the incidence and prevalence of atrial fibrillation is rising worldwide.[29-31] The United States is no exception: analyses of the Framingham Heart Study have shown prevalent and incident AF increasing in the past 50 years in both men and women, and an analysis in Minnesota using ECG data showed the incidence of AF increasing by 12.3% from 1980 to 2000.[32,33] If these trends continue, there will be 15.9 million U.S. adults with AF by 2050 rather than previous estimates of 5.6 million (**Figure 2**).[32] These changes are hypothesized to be due to enhanced surveillance, reduced mortality from a host of risk factors for AF, and higher population life expectancy.[33] This increased prevalence is paired with high economic costs: the annual adjusted per capita medical cost of individuals with AF using U.S. claims data is $10,355 higher than that of individuals without AF for patients 18-65 and $3,600 higher for patients over 65, with

undiagnosed AF only adding to the overall cost to the U.S. healthcare system.[34] By far the most costly cardiovascular outcome of AF is stroke, which imposes a large burden on both the patient and various caregivers. Direct costs to Medicare for AF-related strokes have been estimated at $2.6 billion.[35] As more patients are diagnosed with AF and the population ages, identifying optimal pharmacologic treatments to reduce the cardiovascular AF outcomes of stroke and mortality is key to improving public health and preventing high costs to publicly-funded and private insurers.

2.2: Pharmacologic Treatment Options for Anticoagulation

2.2.1: Warfarin

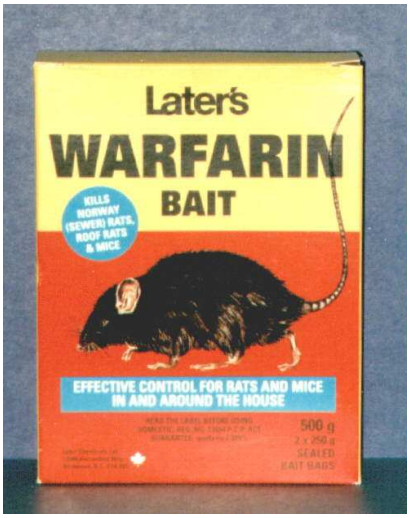

Figure 3: Warfarin as it was available in the early 20th century.

Warfarin (also known as coumadin or coumarin) has been the treatment of choice for oral anticoagulation in patients with AF at high risk for stroke for more than 50 years.[22] Initially used as a rat poison, potential health benefits as an anticoagulant were identified at the University of Wisconsin.[36] Warfarin prevents the synthesis of several key vitamin K dependent clotting factors in the liver. This leads to reduced thrombolysis, stroke, and death. Unfortunately, warfarin features both a narrow therapeutic range and an extremely long and variable half-life between 20 and 60 hours.[37] As a result, medical providers need to carefully titrate and monitor the dosing of individuals on warfarin to ensure their international normalized ratio for clotting (INR) stays within the therapeutic range (generally between 2 and 3 for non-valvular AF patients and slightly higher for patients with prosthetic heart valves) to prevent excessive bleeding events.[37] These INRs are an estimate of the time it takes blood to clot and are typically taken at least weekly while doses are being initially adjusted. Even when a patient is stable with a consistent warfarin dose, guidelines recommend patients are monitored monthly (though extremely consistent and reliable patients may be suitable for monitoring every two months).[22]

Further complicating matters, due to warfarin's metabolism by multiple cytochrome P 450 (CYP) enzymatic pathways and mechanisms of action, serum concentrations of warfarin can be substantially altered by antibiotics, herbal remedies, and even the amount of green vegetables in an individual's diet.[38] This creates a panoply of interactions that can quickly result in subtherapeutic or supratherapeutic INRs unless dosing is carefully monitored, even when warfarin is being taken as directed. While antidotes to warfarin-induced bleeding are straightforward due to the low cost and ease

of administering a vitamin K infusion, long-term damage from bleeding events can be severe and the hospitalizations still have a mean cost of $10,819.[39] While many attempts have been made to identify optimal methods for initiating patients on warfarin using both randomized trials and observational data, including attempts to leverage pharmacogenomics, [40-42] the science of warfarin dosing remains challenging. Still, anticoagulation with warfarin is the definitive benchmark for thromboprophylactic AF treatment and all alternative medications must be able to demonstrate non-inferiority in safety and efficacy relative to warfarin to obtain regulatory agency approval in both the United States and internationally.

2.2.2: Novel Oral Anticoagulants (NOACs)

Due to advances in biological understanding and chemical engineering, the 21st century saw the advent of additional oral anticoagulant options.[36] Instead of targeting vitamin K synthesis to reduce levels of circulating clotting factors, these agents either directly inhibit factor Xa or thrombin.[43] NOACs possess a significantly shorter half-life compared to warfarin and a much wider space where doses are both efficacious and safe, making therapeutic monitoring much less necessary during the course of long-term anticoagulation.[43] Between these characteristics and their smaller suite of drug interactions, there is limited to no need for the patient-specific dosing characteristic of warfarin treatment beyond potential renal dose adjustment.[22] Large-scale randomized controlled trials (RCTs) were performed to assess non-inferiority to warfarin with respect to both safety and efficacy, and the results were sufficiently positive to allow approval and marketing of many NOACs in the United States and elsewhere.[44-47] Since their approval, NOACs have increased in market share in clinical cohorts to the point where warfarin is used in fewer than 50% of patients in some countries, allowing comparative studies of warfarin to the NOACs similar to the one proposed in the current study.[48,49] Notably, NOACs

have generally only been approved for the treatment of non-valvular AF, so valvular AF is still treated with warfarin.

Figure 4: Dabigatran packaging.

2.2.2.1: Dabigatran

Dabigatran (brand name Pradaxa) was the first of the NOACs to be approved in the United States for use in patients with non-valvular AF on October 10th, 2010. Dabigatran is a direct thrombin inhibitor, binding to thrombin to prevent thrombin's conversion of fibrinogen to the fibrin used to bind together platelets and create blood clots. It has a half-life of between 12 and 14 hours that is consistent across patients with normal renal function, resulting in twice-a-day dosing, and is eliminated renally, hepatically, and through P-glycoprotein pumps.[50] Because it is not a CYP substrate, it possesses far fewer drug interactions than warfarin (though it interacts with P-glycoprotein inhibitors like proton-pump inhibitors) and because it acts on thrombin directly its effectiveness is not dependent on the amount of vitamin K a patient consumes. While the RE-LY trial examined both 110 mg and 150 mg dabigatran doses twice daily, only the 150 mg dosage was approved in the United States with a 75 mg dosage available for individuals with creatinine clearance between 15 and 30 mL/min. As of October 2015, Dabigatran is the first of the NOACs to have its own targeted reversal agent, idarucizumab, which binds with dabigatran at the molecular level with higher affinity than thrombin to rapidly reverse dabigatran's effects.[51-53] Boehringer Ingelheim has posted the trial on ClinicalStudyDataRequest.com, a site allowing investigators access to individual-level trial data.

2.2.2.2: Other NOACs

Several other drugs have recently been discovered that fall under the umbrella of novel oral anticoagulants. Rivaroxaban (approved in the U.S. for AF on November 4th, 2011), apixaban (approved in the U.S. for AF on December 28th, 2012), and edoxaban (approved in the U.S. for AF January 8th, 2015) inhibit factor Xa of the clotting cascade, meaning they have a slightly different mechanism of action than dabigatran.[54] Notably, rivaroxaban is the only one of these agents with once-daily dosing for prevention of stroke in AF. Each of these agents was approved after a randomized non-inferiority trial comparing them to warfarin: the ROCKET-AF trial by Bayer for rivaroxaban,[47] the ARISTOTLE trial by Pfizer and Bristol-Myers Squib for apixaban,[44] and the ENGAGE-AF trial by Daiichi Sankyo for edoxaban.[46] In these trials, apixaban and edoxaban demonstrated superiority to warfarin with respect to both stroke prevention and major bleeding, rather than the stroke superiority and major bleeding non-inferiority shown by dabigatran and rivaroxaban. Andexanet alfa has been developed as an antidote for major bleeding associated with the use of any of these Xa inhibitors.[55] Unfortunately, none of these trials are posted on ClinicalStudyDataRequest.com; individual-level or joint categorical trial data would need to be obtained through direct contact with the pharmaceutical company. Their later approval (particularly for apixaban and edoxaban) means that comparative effectiveness studies have a smaller sample size, particularly if data is only available through 2015 or the advent of ICD-10 codes, though there have still been several studies looking at each of them.

2.3: Existing Evidence on Dabigatran Safety and Efficacy

As with any approved drug, a great deal of research has been conducted in efforts to gauge the safety and efficacy of dabigatran. Several randomized controlled trials and multiple observational studies have sought to determine the relative merits of each potential pharmacologic treatment option. **Table 1** at the end of **section 2.3** lists the results of some of these pivotal studies; more specifics regarding methodology and potential reasons for differing results are given throughout **Section 2.3**, as are the results of a phase II randomized controlled trial and subgroup analysis results from RE-LY.

2.3.1: Randomized Controlled Trials

2.3.1.1: RE-LY Trial

Dabigatran's approval for the indication of stroke prophylaxis in AF was based on the RE-LY non-inferiority trial.[6] In RE-LY, more than 18,000 AF patients from 951 clinical centers in more than 44 countries were randomized to receive warfarin under a standard dosing protocol, twice daily dabigatran 110 mg, or twice daily dabigatran 150 mg. Patients were blinded to which dose of dabigatran they received, but use of warfarin was open-label. Patients were followed for a variety of outcomes in an intention-to-treat analysis. Markedly reduced hazards were observed for the primary efficacy outcome of stroke or systolic embolism for patients in the 150 mg dabigatran arm versus the warfarin arm (HR: 0.66, 95% C.I. 0.53, 0.82) and major bleeding overall was also slightly lower (HR: 0.93, 95% C.I. 0.81, 1.07), but much higher rates of gastrointestinal bleeding were observed (HR: 1.50, 95% C.I. 1.19, 1.89). Notably, the HR for ischemic stroke (rather than combined hemorrhagic and ischemic stroke) was 0.76 (95% C.I. 0.60, 98) in the 150 mg arm. The 110 mg dose of dabigatran, on the other hand, had a larger decrease in major bleeding versus warfarin (HR: 0.80, 95% C.I. 0.69, 0.93) and a smaller increase in gastrointestinal bleeding (HR: 1.10, 95% C.I. 0.86, 1.41) but a much smaller improvement in the primary efficacy outcome (HR: 0.91, 95% C.I. 0.74, 1.11). Both the 110 mg and 150 mg dosage improved survival

versus warfarin with a HR of 0.91 (95% C.I. 0.80, 1.03) and 0.88 (95% C.I. 0.77, 1.00), respectively. The non-inferiority margins set by the investigators were all met, and the drug was approved for dosing at 150 mg twice daily. It is not clear, however, how these tradeoffs might manifest in populations with differing distributions of risk factors for bleeding and stroke.

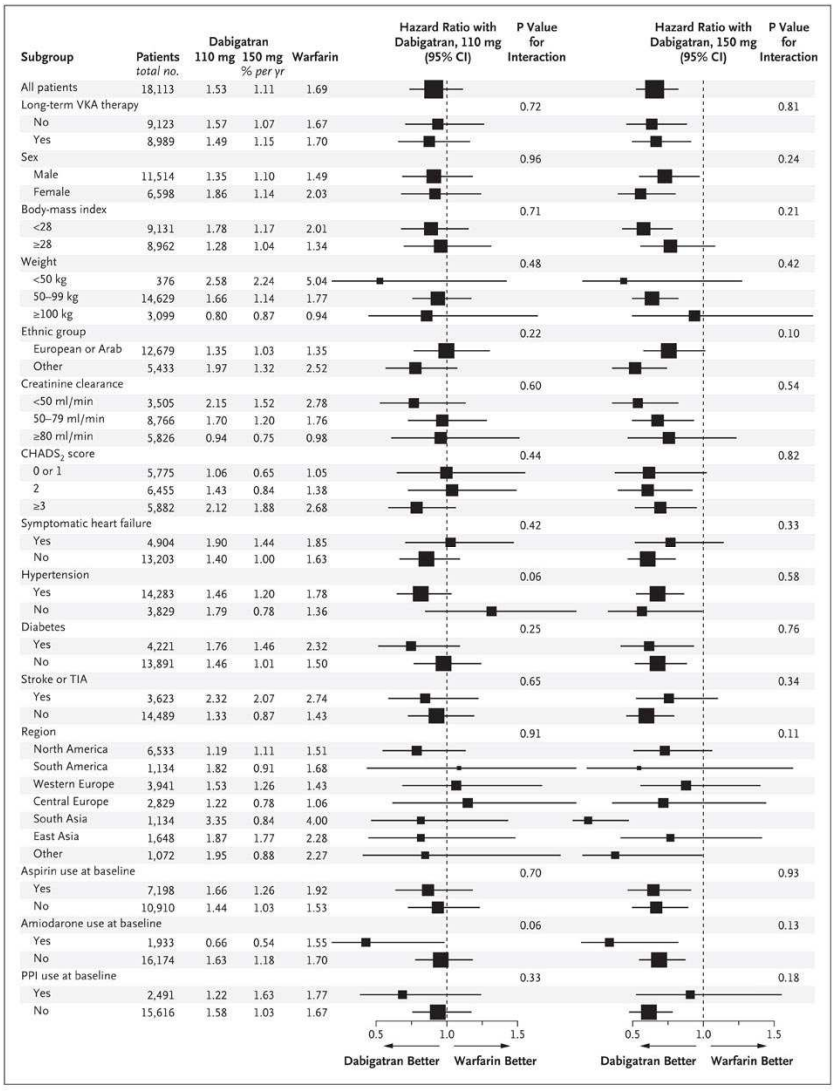

Figure 5: Sub-group treatment effect estimates from the RE-LY trial.

2.3.1.2: Evidence for Heterogeneity in RE-LY

Some trialists have taken the position that subgroup analyses (and effect heterogeneity) should be ignored and the population relative effect should simply be applied to the estimated risks of individuals in each subgroup.[56,57] This represents a significant assumption that effectively ignores the potential for biological differences in dose or individual metabolism of the studied drug. There is some evidence for heterogeneity of dabigatran effect in the RE-LY trial on the multiplicative scale used for the main analyses ($p < 0.25$ for the interaction) with respect to ethnic group, amiodarone use, sex, body-

mass index, and PPI use at baseline, as seen in **Figure 5**. Furthermore, age was not examined in their efficacy subgroup analyses. All of these factors may differ between the population of RE-LY patients and the Medicare AF patients; see **Section D.1** for additional details. In addition, if the other variables in the table are associated with AF outcomes and are indeed not effect modifiers on the multiplicative scale, they will be effect modifiers on the absolute scale and necessary for estimating a valid risk difference and number needed to treat (NNT) for the clinical cohort.[58] Furthermore, subsequent analyses of the RE-LY trial specifically analyzing different types of bleeding risk demonstrated relative scale effect heterogeneity based upon age when categorized above and below 75, particularly for extracranial hemorrhages.[59,60] Similar subgroup analyses were conducted for the stroke outcome in patients with heart failure,[61] patients using other antiplatelet therapies,[62] patients with history of stroke or transient ischemic attack,[61] and patients with paroxysmal vs persistent vs intermittent AF,[63] with none of the analyses finding substantial heterogeneity on the hazard ratio scale but typically preserving benefit.

2.3.1.3: Other Trials of Dabigatran vs Warfarin

There were two additional relevant trials for dabigatran that investigated its safety or efficacy, both showing roughly similar results to RE-LY: The PETRO trial and the RE-COVER trial. The PETRO trial was particularly relevant as the first trial evaluating dabigatran in patients with AF and showed similar bleeding risk between 150 mg of dabigatran twice daily and warfarin in the AF population.[64] As a phase II trial, it was much smaller than the RE-LY trial, enrolling 502 patients and randomizing them to one of four trial arms: 50 mg of dabigatran twice daily, 150 mg of dabigatran twice daily, 300 mg of dabigatran twice daily, or warfarin with a target INR of 2-3, with the warfarin arm being open-label. If patients were randomized to one of the dabigatran groups, they were further randomized to take no aspirin, 81 mg of aspirin, or 325 mg of aspirin once daily as well. Like RE-LY, PETRO enrolled individuals with AF that were at high risk of thromboembolic events. Like RE-LY, PETRO was conducted in international study centers,

though the study centers were generally in European countries as well as the United States. Unlike RE-LY, PETRO was a phase II study focused on determining a safe dose of dabigatran that would result in an acceptable level of bleeding events relative to warfarin. It also sought to identify efficacy by examining the anticoagulant activity biomarkers of activated partial thromboplastin time and inhibition of D-dimer generation. Study investigators did not test a formal statistical hypothesis, instead seeking mainly to measure data on pharmacodynamics to help decide which doses should be used in the RE-LY study. Major bleeding events only occurred in the 300 mg of dabigatran twice daily group taking 81 mg or 325 mg of aspirin, and there were generally more bleeding events with higher doses of dabigatran and higher doses of aspirin. Of the groups, only the 50 mg dabigatran group had substantially lower incidences of bleed than warfarin; the 150 mg dabigatran group was comparable. Only the 50 mg group experienced any embolic events. Biomarker analyses bore out these results. The end result of the study was to move forward with the 150 mg twice daily dabigatran dose, rather than the 50 mg or 300 mg dose, due to its favorable bleeding and embolic profile.

The second study, the RE-COVER trial, was an investigation of the use of dabigatran when treating venous thromboembolism and showed some potential heterogeneity in bleeding effects.[65] RE-COVER enrolled 2,564 patients with acute venous thromboembolism from 228 clinical centers in 23 countries and randomized them to receive dabigatran 150 mg twice daily or warfarin with a target INR of 2-3. RE-COVER assessed both safety and efficacy in the setting of venous thromboembolism, and since it used a similar dosage of dabigatran as RE-LY the safety information is particularly valuable. Unlike RE-LY, however, RE-COVER was a double-blind double-dummy trial, where individuals taking dabigatran were given placebo warfarin pills and even had placebo INRs and dose adjustments performed, and it had a shorter follow-up time (six months). Unlike the results of RE-LY, RE-COVER found strong reductions in major and clinically relevant bleeding for dabigatran 150 mg twice daily (HR: 0.63, 95% C.I. 0.47, 0.84) and for any bleeding event (HR: 0.71, 95% C.I. 0.59, 0.85), though there were

more gastrointestinal bleeds in the dabigatran group than the warfarin group (53 vs. 35). Because RE-LY focused on including individuals with cardiovascular risk factors, including age and diabetes, RE-COVER's population was generally younger (mean age 71.5 in RE-LY vs mean age 55.0 in RE-COVER), suggesting there may be some modification of bleeding effect by age or these other risk factors that could account for the improved overall bleeding profile in RE-COVER.

2.3.2: Non-experimental Studies

There have been several comparative effectiveness studies published using routintely collected claims data in an attempt to determine whether dabigatran and other NOACs are as or more effective in practice compared to their performance in clinical trials. Their results have varied, but all have differed from RE-LY in one way or another. On the whole they reinforce the need for additional studies and examination of potential treatment effect heterogeneity.

Perhaps the largest study to date has been the FDA's analysis using initiators of warfarin and dabigatran in the full Medicare sample from October 2010 to December 2012.[15] They identified 67,494 new initiators of dabigatran and 273,920 new initiators of warfarin including individuals with any inpatient or outpatient diagnoses for AF or atrial flutter and excluding individuals with other competing indications for warfarin treatment. They used propensity score matching to eliminate differences in their study variables; based upon standardized mean differences (SMDs) they were successful and all dabigatran individuals had a match. They used an as-treated design, censoring individuals from the analysis after switching anticoagulants or a gap in days' supply of 3 days, though a sensitivity analysis examined the ramifications of allowing a grace period of up to 14 days. Their death outcome was not actually death; instead, they analyzed deaths not preceded by a study outcome or within 30 days of hospitalization for a study outcome, making it difficult to interpret. They estimated a hazard ratio of 0.80 (95% C.I. 0.67, 0.96) for ischemic stroke, favoring dabigatran, and a hazard ratio of 0.97 (95% C.I.

0.88, 1.07) for major bleeding, with a heightened risk of gastrointestinal bleeding for dabigatran patients with a hazard ratio of 1.28 (95% C.I. 1.14, 1.44); when they looked at dosages of 150 mg specifically, rather than 75 mg, they found improved stroke reduction (hazard ratio of 0.70) but greater risk of gastrointestinal bleed (hazard ratio of 1.51) more in line with the results of the RE-LY trial. They did notice, however, that the increase in the risk of gastrointestinal bleeding was concentrated in older adults, particularly women over 75 and men over 85. Women over 85 were the only group with an increased risk of mortality from dabigatran, but other groups experienced an improvement in mortality with an overall hazard ratio of 0.76 (95% C.I. 0.67, 0.86) with the 150 mg dose, more pronounced than the 0.88 observed in the RE-LY trial. The fact that they implemented an as-treated design and used such a short gap in days supply despite the large variance in warfarin prescribing makes it difficult to interpret the extent to which their findings parallel those from RE-LY. Still, the results were largely similar to the trial results on the relative scale, with a lower hazard ratio for mortality and some evidence of bleeding effect heterogeneity based on age and sex.

At the same time the FDA study was running on the full Medicare sample, Hernandez et al were investigating the question using the 5% Medicare sample.[16] This study only used Medicare initiators from October 2010 to October 2011, limiting the population, and required two outpatient diagnoses for AF or atrial flutter to qualify an individual as an AF patient, as well as requiring individuals to fill within two months of their incident diagnosis (while the FDA merely required a diagnosis at any time before filling). The resulting cohort had only 1,302 Dabigatran users and 8,102 warfarin initiators, who they compared after using inverse probability of treatment weighting based on a propensity score. They also used an as-treated design but provided a 60 day grace period and censored individuals at death. They focused predominantly on bleeding outcomes, rather than stroke or systemic embolism. They provide p values rather than SMDs so it is difficult to determine how successful their weighting process was, but it seems to have performed well for most variables. While it would be quite useful to see the extent of the

propensity score overlap to compare the difference between the dabigatran user and whole population, they do not share this information in their paper or supplement. After adjustment, they identified an increased risk of major bleeding with a hazard ratio of 1.58 (95% C.I. 1.36, 1.83) and an increased risk of gastrointestinal bleeding with a hazard ratio of 1.85 (95% C.I. 1.64, 2.07), both higher than the original RE-LY trial. Despite being conducted on a subsample of the FDA data, their results looked quite a bit worse for dabigatran, suggesting that either the difference in grace period, the use of a different target population than the FDA (the whole population of initiators rather than the dabigatran patients targeted in a matched design), worse confounding control, or all three are to blame.

Other U.S. governmental databases have been used to answer this question as well. Villines et al utilized the Department of Defense database, which provides uniform medical coverage and pharmacy benefits to nearly 10 million individuals receiving care at both military and non-military institutions, to find initiators of dabigatran and warfarin from October 2010 to July 2012 with at least one diagnosis for AF within 12 months of their initiation.[17] They identified a cohort of 14,813 dabigatran users and 24,500 warfarin initiators, with propensity score matching using a model built by backwards selection reducing it to 12,793 of each. These authors used a grace period of 30 days after the final prescription, with sensitivity analyses exploring the impact of a 45 and 60 day grace period and integrating INR measurements as a way to extend follow-up, though it is unclear how they identified INR measurements. They stopped following patients at death. Comparing dabigatran to warfarin, they estimated a hazard ratio for stroke of 0.73 (95% C.I. 0.55, 0.97), a hazard ratio for major bleeding of 0.87 (95% C.I. 0.74, 1.03) and 0.82 (95% C.I. 0.71, 0.95) when restricting to 150 mg doses, and a hazard ratio for gastrointestinal bleeding of 1.13 (95% C.I. 0.94, 1.37). Additionally, they found a hazard ratio for myocardial infarction of 0.65 (95% C.I. 0.45, 0.95), much lower than that observed in RE-LY of 1.35 (95% C.I. 0.98, 1.87); similarly, the hazard ratio for death was 0.64 (95% C.I. 0.55, 0.74). The authors state none of their sensitivity analyses examining the impact of longer or INR-supplemented follow-up periods

changed results much. Overall, their results were more similar to the RE-LY results than the other non-experimental studies, but the fact that the gastrointestinal bleeding, M.I. and death estimates were substantially better for dabigatran than those from RE-LY is notable. Unfortunately, it is impossible to determine whether these differences are because of poor confounding control or heterogeneity, measured or otherwise, between this and the RE-LY trial population.

Other researchers examined private insurance databases. Seeger, Schneeweiss, et al performed an analysis using data from two commercial insurance databases (MarketScan from Truven and Clinformatics from Optum) focusing on both safety and efficacy.[18] These investigators identified 41,103 warfarin and 18,560 dabigatran initiators between October 2010 and December 2013. The authors matched on a propensity score with 78 investigator-specified covariates, resulting in a final cohort of 15,529 initiators of each medication (with successful removal of imbalances in the covariates as measured by the SMD), with an additional analysis using high-dimensional propensity scores. They performed an as-treated analysis with a grace period of 14 days without any days supply on hand, but also performed a sensitivity analysis extending this period by 365 days in an imitation of an intention-to-treat analysis. This study was the first to look at benefits on an absolute scale, rather than reporting purely hazard ratios. Individuals stopped follow-up at the time of death. Unlike many of the other studies, they combined ischemic and hemorrhagic strokes in their analysis. When contrasting dabigatran with warfarin, they found a hazard ratio of 0.77 (95% C.I. 0.54, 1.09) and one-year risk difference of -0.0003 (95% C.I. -0.0006, 0.0002) for strokes and a hazard ratio of 0.75 (95% C.I. 0.65, 0.87) and risk difference of -0.018 (95% C.I. -0.025, -0.010) for major hemorrhages; they did not investigate gastrointestinal hemorrhages or mortality. P values for the proportional hazards assumption were 0.23 and 0.18 for stroke and major bleeding. Despite their more pronounced hazard ratio for major bleed compared with the trial (0.75 vs. 0.93), they did not detect any “significant” heterogeneity in major bleeding across their subgroups. On the other hand, they noted some trends towards improved

dabigatran stroke reduction in adults over 75 and improved reduction in major hemorrhage in adults under 55. All of this suggests that an older population (such the warfarin users in their analysis) might experience a differing risk/benefit profile than that in this study on both the absolute and relative scale.

Another large-scale study was conducted in Denmark, where both the 110 mg and 150 mg twice daily doses of dabigatran were approved for use in August 2011.[66] They included dabigatran initiators after August 2011 and only allowed warfarin initiators to enter between August 2009 and July 2010, resulting in initial cohorts of 5,106 dabigatran and 13,548 warfarin patients. This was another propensity-score matched analysis (this time matched 2:1) that found matches for 4,978 dabigatran patients of either dosage and censored at treatment switching but not discontinuation with two levels of propensity score; they built their model for treatment choice using warfarin initiators after August 2011 and then applied it to warfarin users from 2009 to 2010, and also built a model predicting which type of dabigatran an individual would initiate if they did initiate dabigatran. This latter step was designed to allow them to perform contrasts for each of the doses, but did not appear successful: quite a few of their covariates show SMDs greater than 10% for one or both of the dosages, including key confounders like renal function, sex, age, and some of the medication usages at baseline. Possibly as a result, they found large mortality benefits for both doses of dabigatran with hazard ratios of 0.79 (95% C.I. 0.65, 0.95) for the 110 mg dose and 0.57 (95% C.I. 0.40, 0.80) for the 150 mg dose compared to warfarin. They also found much lower risks of gastrointestinal bleeding comparing dabigatran 110 mg with warfarin with a hazard ratio of 0.60 (95% C.I. 0.37, 0.93) and only a slight increase for 150 mg with a hazard ratio of 1.12 (95% C.I. 0.67, 1.83), with favorable results for major bleeding with both dosages as well (HR: 0.82 (95% C.I. 0.59, 1.12) for the 110 mg dose and HR: 0.77 (95% C.I. 0.51, 1.13) for the 150 mg dose). Stroke benefits were also inconsistent with findings from other studies and with the logical assumption that higher doses will prevent more embolisms, with hazard ratios of 0.73 (95% C.I. 0.53, 1.00) for the 110 mg dose and 1.18 (95% C.I. 0.85, 1.64) for the 150 mg dose. Given the large differences

between their cohorts even after propensity score matching and the strange trends between dabigatran doses (which could be the result of matching to different cohorts), their results are difficult to interpret but seem to show more pronounced benefits for dabigatran compared to the results from the RE-LY trial and other non-experimental studies.

The final and most recent study focusing specifically on the contrast between dabigatran and warfarin in AF was conducted by the FDA with the Sentinel system in 2017.[20] The Sentinel network collects data from a variety of administrative, clinical, and pharmacy dispensing databases for use in large-scale investigations of key medical questions in the United States.[67] Using this system and data from November 2010 to May 2014, Go et al conducted an propensity-score matched analysis (with matching and model estimation performed within each of the data partners), identifying 25,289 dabigatran initiators and finding matches for each one amongst the 83,034 warfarin initiators. They conducted an as-treated analysis with a grace period of 7 days between prescriptions and allowance for stockpiling on medications. Individuals were censored at death. Because they used many claims-based insurance sources, their mean age of 68.4 was lower than many of the Medicare-based studies or the RE-LY trial itself. Their propensity score was quite successful in removing the differences between all their measured covariates (as one might expect with such a large sample). They estimated a hazard ratio for ischemic stroke of 0.92 (95% C.I. 0.65, 1.28) and a hazard ratio for gastrointestinal bleeding of 1.04 (95% C.I. 0.83-1.30), meaning their results showed both less benefit and less harm than the RE-LY trial. Interestingly, however, they did identify substantial heterogeneity in gastrointestinal bleeding with those under 65 having a hazard ratio of 0.59 (95% C.I. 0.32-1.07), those between 65 and 74 having a hazard ratio of 0.81 (95% C.I. 0.52-1.24), those between 75 and 84 having a hazard ratio of 1.47 (95% C.I. 1.05, 2.14) and those over 85 experiencing a hazard ratio of 1.84 (95% C.I. 1.05, 3.20). There was also some heterogeneity by kidney function in both ischemic stroke and gastrointestinal bleeding risks, though this may be due to other concomitant factors associated with reduced kidney function like age,

hypertension, diabetes, and congestive heart failure. Overall, while they used an as-treated design that differed from RE-LY and included a younger patient population, these results reinforce that the overall benefits and risks of treatment with dabigatran may be quite heterogeneous across populations and that these heterogeneities in risk may only become apparent in large databases.

There is another study focusing on several different NOACs with dabigatran as one of the potential options whose results also warrant discussion. Lip et al conducted a study in MarketScan examining NOAC initiators from January to December 2013 focusing specifically on major bleeding risk and using Cox proportional hazards with direct adjustment for a variety of variables and backwards selection at $p < 0.2$.[11] These authors identified a decreased rate of major bleeding for dabigatran relative to warfarin (HR: 0.88, 95% C.I. 0.64-1.21). Their propensity matched analysis showed a slightly reduced rate of major bleeding with a hazard ratio of 0.69 (95% C.I. 0.50-0.96), suggesting some potential treatment effect heterogeneity.[12] In both analyses individuals were followed until discontinuation from their initial medication or switching, though the amount of gap or grace period they allowed is unclear. These results generally seem to agree with those of other studies in younger, claims-based cohorts.

Overall, there is still significant clinical uncertainty about the actual benefit/risk profile of dabigatran compared to warfarin; still, these studies have clearly shed light on the fact that there is likely heterogeneity in the safety and efficacy of dabigatran with respect to warfarin. While each of the studies added to our body of knowledge only one of them implemented an intention-to-treat analysis for comparison with RE-LY. Beyond this, almost all of the studies focused on estimating a treatment effect amongst dabigatran initiators who have characteristics that set them apart from the general patients included in RE-LY. Most stopped following individuals at their time of death while using Cox proportional hazards and Kaplan-Meier estimators that assume we can prevent competing risks; this can

be problematic, particularly when the competing event is roughly as common or more common than the event of interest.[68] A new study in a Medicare cohort spanning more time using both IPTW and matched designs alongside an analysis using the trial data that conditions on staying on therapy would improve understanding of which medication performs better in an older population with a higher prevalence of comorbid conditions without assuming adherence patterns will be constant across the RE-LY and Medicare cohorts with respect to baseline modifiers.

Table 1: Results of published studies estimating the safety and effectiveness of dabigatran with respect to warfarin.

Study population	Type	Ischemic Stroke/Embolic Event Results	Gastrointestinal Bleeding Results	Mortality Results
RE-LY[6]	ITT RCT	HR: 0.76 (0.60-0.92)	HR: 1.50 (1.19-1.89)	HR: 0.88 (0.77-1.00)
RE-COVER[65]	ITT RCT	------	53 bleeds (dabigatran) vs 35 bleeds	------
100% Medicare sample[15]	As-treated propensity-matched*	HR: 0.80 (0.67-0.96)	HR: 1.28 (1.14-1.44)	HR: 0.76 (0.67-0.86)
5% Medicare sample[16]	As-treated IPTW	-------	HR: 1.85 (1.64-2.07)	------
DOD database[17]	As-treated propensity-matched*	HR: 0.73 (0.55-0.97)	HR: 1.13 (0.94-1.37)	HR: 0.64 (0.55-0.74)
U.S. commercial claims[18]	As-treated propensity-matched*	HR: 0.77 (0.54-1.09)	------	------
Denmark[19]	As-treated propensity-matched*	HR: 1.18 (0.85-1.64)	HR: 1.12 (0.67-1.83)	HR: 0.57 (0.40-0.80)
Sentinel database[20]	As-treated propensity-matched*	HR: 0.92 (0.65-1.28)	HR: 1.04 (0.83-1.30)	------

*Matched to the cohort of dabigatran initiators

2.4: Heterogeneity Between Trial and Clinical Cohorts

RCTs have long been held up as the gold standard for assessing the effects of various interventions, and for good reason.[58] Randomization of patients will in expectation prevent confounding of effect estimates and often allows for straightforward data analysis. Unfortunately, the non-random nature of trial inclusion can lead to differences between trial cohorts and source populations in key variables, limiting our ability to directly transport results from the trial to target populations.[69] There are three main types of heterogeneity that are concerning when attempting to transport the estimates of the RE-LY trial: demographic characteristic heterogeneity, warfarin management heterogeneity, and medication adherence heterogeneity.

2.4.1: Demographic Characteristic Heterogeneity

Heterogeneity in demographic characteristics between trials and source populations is a common concern in the field of cardiology. A systematic review concluded that patients of advanced age and patients with complex comorbidity profiles are often not selected into randomized controlled trials in cardiovascular trials.[10] Some studies have compared distributions of age, sex, race, and other baseline characteristics in trial patients compared to large-scale population estimates from stroke registries and found that trial populations are generally younger and more male.[9] This is especially true when comparing to older populations of patients initiating dabigatran. For example, in the FDA study using Medicare, the mean age was more than 75 compared to the RE-LY trial's mean age of 71, 51% of initiators were female rather than RE-LY's 64% male, and 33% of patients had diabetes rather than the 23% in RE-LY.[15] Other researchers have assessed the proportion of patients seen in their clinic which would have been eligible for trial enrollment, finding that less than half of patients in a suspected stroke registry with AF were eligible for inclusion into the trials for the direct oral anticoagulants.[7] The overwhelming consensus is that trial samples are not a representative sample of the clinic population. If

any of the differing factors amongst eligible patients alter the treatment effect observed in the trial, adjustment of effect estimates will be necessary to obtain an unbiased effect estimate in the eligible patients.

2.4.2: Warfarin Management Heterogeneity

An additional source of heterogeneity between trials and clinical populations unique to oral anticoagulation is differences in methods of initiating and altering dosages of warfarin between trial and clinical practice.[21] For example, RE-LY's suggested warfarin management protocol involved specific changes in warfarin management with INRs observed in specific ranges; variance from this protocol was associated with lower time in therapeutic range (TTR, the amount of time patients spent with a safe and effective INR and a common metric for evaluating warfarin management quality).[70,71] As a result of these algorithms and potential for better care, TTRs observed in trial patients can vary compared to the TTRs observed in community practice.[71-73] While the median TTR in the RE-LY trial across all centers was 66%, national estimates for TTRs in the United States are between 54% and 55%.[74,75]

Outside trials, TTRs can vary depending on the type of provider managing anticoagulation, with cardiology clinics generally having higher TTRs than primary care at 61% vs 55% with only 10% of their time supratherapeutic rather than 15%.[76] One analysis based in active anticoagulation clinics testing new methods for predicting effective warfarin management even showed TTRs above 75%.[77] Even within the RE-LY trial, there was still substantial variation in INR control across the various centers that may have contributed to patients experiencing more or less benefit compared to being randomized to dabigatran.[78] This is in part due to difficult to manage patients being excluded from trials, but different management methods are also likely to contribute. Dabigatran management and dosing is straightforward; it simply requires twice daily dosing with no target therapeutic range with some dosage adjustment based upon renal function, resulting in much better consistency than that observed with

warfarin. With all of these differences in warfarin management between the clinical trial and routine care setting, it becomes increasingly important to check the assumptions for generalizability when data does not exist to include warfarin management quality in the sampling model.

2.4.3: Medication Adherence and Persistence Heterogeneity

Adherence and persistence must also be considered in addition to management strategy. While estimating intention-to-treat effects ordinarily allows researchers to ignore adherence in the context of randomized controlled trials, if individuals outside the trial differ in their adherence to those included in the trial bias can be introduced into a transported effect estimate.[79] Unfortunately, participation in a clinical trial also generally improves both adherence and persistence.[80]

If adherence were not linked closely linked to outcomes with the two medications, this problem might be ignorable. Regrettably, adherence is a clear cause of heterogeneity for both NOACs and warfarin; poor adherence to NOACS may be associated with short-term spikes in stroke risk and warfarin has a long history linking missed doses to changes in INR and treatment effects.[81] To make matters worse, the effects of trial participation may be differential depending on whether patients are taking warfarin or a NOAC, particularly if trials use placebo INR visits in the NOAC arm or see patients more frequently than is expected in clinical care (which is less likely to be the case with warfarin, which requires office visits even when managed well). Dabigatran adherence may also differ between trial and clinical populations due to the price of the medication. Studies conducted after the introduction of NOACs have generally shown much higher levels of non-persistence than in clinical trials, with up to 60% of individuals with AF in commercial insurance in the United States discontinuing warfarin after 1.1 years and more than 50% of patients discontinuing NOACs, compared to RE-LY's 16% and 10% at one year.[82] Similar increases in NOAC and warfarin discontinuation compared to RE-LY were observed in Danish, Canadian, and UK populations, though they were not as extreme.[83-85] Using methods to assess whether

patients in the trial are experiencing drastically different outcomes from individuals in the target population after controlling for known modifiers can help elucidate the degree to which treatment effect estimates are influenced by adherence and management concerns, as can making sure both trial and target population estimates dealing with populations with similar adherence.

2.5: Methods for Transporting Causal Effects

Multiple methods exist for transporting and generalizing treatment effects from studies to differing populations. To give some perspective on the reasoning behind the methods that will be used in this study, we present here a brief summary of the oldest method for transporting results (direct standardization) as well as the advantages weighting methods have.

2.5.1: Direct Standardization Methods

Standardization is a potent and necessary tool for comparing results across different populations, whether to assure internal validity and remove confounding or to improve external validity and remove problems of effect measure modification. As such, standardization has been used in epidemiology for quite some time; the first examples involving standardizing rates of death can be traced back to the 18th century comparing observed to expected survival amongst various differing professions based upon age. It wasn't until the 1970s that Miettinen published a paper describing the potential benefits of standardizing rate ratios to entire referent populations rather than using SMR weighting.[86]

When dealing with a few levels of key categorical variables, standardization is straightforward and simple cases are presented in textbooks[58] describing generalization or transportation of causal effects. In standardization, a specific standard is chosen that identifies the distribution of key strata in a population. Estimates of treatment effects within those key strata are calculated and the results averaged using weights identified from the standard. Creating age category-sex standards is commonplace when performing some epidemiologic calculations, but building three or four-variable standard becomes complex, as does attempting to incorporate continuous variables into the standard. When transporting results to target populations, we often need to deal with a large number of effect measure modifiers that may or may not change how one another modify the treatment effect, making

standardization incredibly cumbersome and potentially imprecise. Moreover, dealing with continuous variables improperly when standardizing (for example, by dichotomizing age) may misrepresent the modification of the treatment effect of exposure and cause bias in the resulting estimates of treatment effects, just as it can result in bias when attempting to control for confounding.

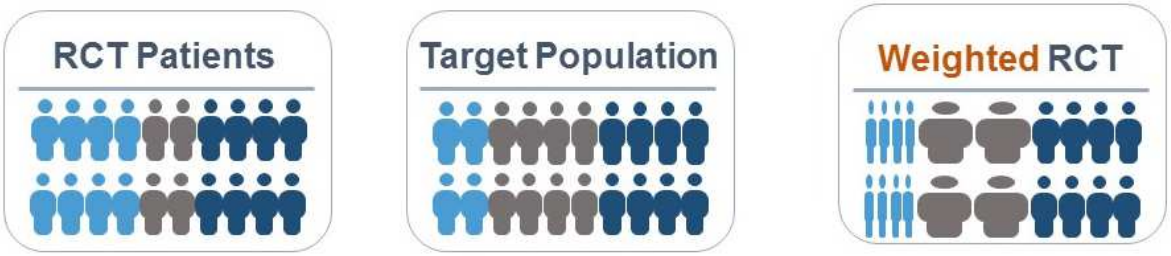

Figure 6. A graphic depicting the results of weighting. Because light blue patients were more likely to be included in the trial and gray patients were less likely, the light blue patients received lower weights while gray patients received higher weights. This means that light blue patients will be less influential in the final analysis of the trial than in the crude analysis.

2.5.2: Weighting Methods

Weighting methods are a more flexible alternative to standardization when generalizing and transporting treatment effects, though they rely on more parametric assumptions.[87,88] In weighting methods, the trial population is transformed into a pseudo-population that resembles the target with respect to selected variables and then re-analyzed using these weights (**Figure 6**). In the setting of small numbers of categorical variables that may be potential effect measure modifiers and fully saturated non-parametric weighting models, weighting methods and standardization yield identical results. With larger numbers of effect measure modifiers or continuous effect measure modifiers, however, weighting methods allow researchers to obtain treatment effect estimates under some parametric assumptions that are cumbersome to implement in standardization.

The actual implementation of weighting methods is fairly straightforward. Rather than creating a standard and computing a weighted average of stratum-specific estimates based on that standard, researchers obtain data on the selected effect measure modifiers in both the trial or study population and target population, preferably at an individual level or tabular level. Aggregate information can also

be used, but this approach ignores the correlation of variables in the target population that could lead to interactions altering treatment effect. In order to transport both absolute and relative scale treatment effects (and thus transport the estimated risks), variables should be included in the sampling model that separate sampling and the outcome on a causal diagram.[88,89]

Logistic regression can then be used to estimate the probability of being in the study population rather than the target population, conditional on the key effect measure modifiers with as few or as many interaction terms as are necessary. From this probability, researchers can construct either inverse probability of sampling weights (IPSW) or inverse odds of sampling weights (IOSW) based on whether or not they wish to consider the study population part of the target population and generalize treatment results (which requires IPSW) or whether they consider the target population its own discrete to which the study results should be transported (which requires IOSW). These weights can be further stabilized based upon the probability or odds of sampling ignoring the key effect measure modifiers such that the sum of the weights will be equal to the original number of trial participants. Mathematically, stabilized odds weights are calculated for individuals in the study population as follows (where "EMM" refers to effect measure modifiers on the scale of interest):

$$Stabilized\ inverse\ odds\ of\ selection\ weight\ (sIOSW) =$$

$$\left(\frac{Odds(study\ population)}{Odds(study\ population \mid EMM)}\right) =$$

$$\left(\frac{P(study\ population)}{1 - P(study\ population)}\right) * \left(\frac{1 - P(study\ population \mid EMM)}{P(study\ population \mid EMM)}\right)$$

With individuals in the target population receiving weights of 0 so that they do not contribute to the analysis. Analyses of interest are then carried out on the trial data using the IOSW or IPSW weights. Notably, variance in estimates resulting from sampling of the target population will be ignored by normal variance estimators or sandwich estimators with bootstrapping required to get conservative standard errors.[87]

2.6: Methods for Assessing Transportability

While these data transport methods exist and have existed in some form for quite some time, they have rarely been utilized. Part of this may be because of the lack of access to trial data, but some of it is surely due to difficulty in knowing when exactly they are necessary as well as how likely they are to actually work. After all, reweighting the trial population results in increased uncertainty that may lead to "non-significant" results, even if there is still evidence for some degree of causal effect. Techniques have been used to assess when we can't transport trial results to a given target population, but they generally ignore the potential for effective use of a trial sampling model. Fortunately, an additional method has been theorized that helps assess performance of these weighting techniques.

2.6.1: Comparing Marginal Variable Distributions

Perhaps the most common method for assessing whether trial results are readily transported or generalized to "real-world" populations is simply comparing the proportion of individuals with a characteristic to the amounts in the real-world, akin to what we discussed in **Section 2.4.1** describing demographic characteristic heterogeneity. These methods are at the heart of the debate around representativeness in randomized controlled trials.[90,91] How does the age distribution in RE-LY compare to that of the general AF population, or the Medicare AF population? Are there more or fewer African Americans than in the general AF population in the United States? Do more or fewer of them have comorbidities like hypertension, diabetes, and other stroke risk factors?

While these questions are all relevant, this metric is imperfect for assessing the overall external validity of trial results. First, if the variable in question is completely unrelated to the outcome, it cannot act as an effect measure modifier and thus will not need to be accounted for. Even if blue-eyed people were underrepresented in a study due to either quirks of enrollment or intensive recruitment, fewer

blue-eyed people in the study will not be a problem if the outcome is stroke and thus unlinked to eye color.

Second, even if the variable is associated with the outcome, its paths may go through other variables that were balanced in the study and target populations; the amount of bias they induce is proportional to the degree to which their association with the outcome remains. Take the issue of age, where trial participants are on balance younger than the general population but age was only associated with stroke because of diabetes. If the prevalence of diabetes were the same between the trial and general population, age would no longer be relevant.

Third, both standardization and weighting can be used to deal with variables that differ between the trial and target population, provided there are at least some individuals included with the characteristic (though uncertainty increases and there is an attendant loss of precision). In other words, marginal imbalance or lack of representativeness is not in and of itself an obstacle to the theoretical exchangeability of the trial and target population.[92] It does, however, represents an indicator for when we may want to standardize or weight trial results. Overall, while this method can indicate when the unweighted trial estimate is likely to be biased, a difference in distribution for a variable alone does not necessarily mean that obtaining an externally valid estimate is impossible, and indeed may be helpful for transporting trial results to a given population.[93]

2.6.2: Assessing Proportion of Patients Eligible for Trials

Another common metric to assess transportability, and one with potentially more merit, is the proportion of patients in a routine care practice or other target population that could have been included in the trial based upon its eligibility criteria. This is essentially the extreme version of examining the marginal distributions of variables, simply looking at whether the prevalence of individuals with the characteristic in the target population is over 0% while it was 0% in the trial. As a result, the same

caveats apply as when comparing marginal distributions; if blue-eyed people were excluded completely from the trial, it is only relevant if being blue-eyed is associated with the outcome even after other variables have been accounted for.

Unfortunately, this is very likely the case with many trial eligibility criteria. They frequently increase the expected number of events to improve power with lower sample sizes or decrease the risk of negative safety outcomes for patient protection. We cannot rely upon weighting and standardization by this discordant variable because lack of positivity makes rendering the trial and target population exchangeable impossible; the trial simply contains no information on individuals with that covariate pattern.[88] Extrapolating past the eligibility criteria requires reliance on non-experimental effect estimates and "generalizability bias," a heavily limited concept dependent on similar confounding structure for eligible and ineligible individuals,[94] or the use of various exposure-variable modeling techniques if the eligibility criteria is a continuous variable. Because both of these methods have quite strong assumptions, a greater the share of the target population ineligible for the trial generally increases the potential for bias in transported effect estimates. Restricting the target population to trial-eligible individuals as much as possible reduces the risk of this kind of bias substantially.

2.6.3: Weighting Methods for Assessing Transportability

There is another potential option that has been underutilized in epidemiology. Transport weights can identify whether outcomes are discordant between trial and target populations after utilizing sampling models. This method was first used in the transport literature by Stuart et al

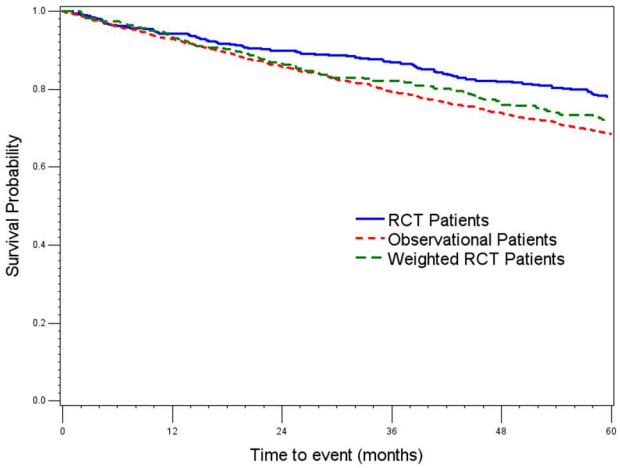

Figure 7: An example where the transport weights seems to work well until 30 months. Survival of the weighted RCT patients is much closer to the observational patients than the unweighted RCT.

using propensity scores in the setting of educational achievement[95] and is analogous to the placebo-checking perspectives used in some trial indirect comparison literature.[96,97] Factors associated with the outcome of interest will be effect modifiers on the absolute scale, relative scale, or both if there is a non-null treatment effect.[58] Thus, a weight specification model that accounts for all heterogeneity on both scales and is properly specified should predict the risk of the outcome in patients in the target population receiving a treatment when it is applied to the patients in the trial receiving that same treatment (in other words, sampling and outcome will be independent on the transport diagram, a requirement of the weighting methods to yield unbiased results).[88] You can then compare the observed target survival curves to the trial survival curves with or without weighting visually (**Figure 7**), at specific time points, or using proportional hazards regression. If there are large residual differences after weighting, the sampling model is inadequate for fully transporting results on both scales. Because we are focusing on the risk differences with respect to ischemic stroke and gastrointestinal bleeding, we can examine their incidence at specified time points in the unweighted trial, weighted trial, and Medicare cohorts to assess if additional effect measure modifiers or consistency violations due to warfarin management or medication adherence makes our transported effect estimates likely to be biased.

Importantly, this approach can only be implemented in one treatment group at a time; it can thus only assess transportability to the treated or untreated, not both at the same time. Because it can only be used on one treatment arm at a time, it is best thought of as a way to rule out the ability to transport treatment effects. Attempts to weight to the entirety of the target population essentially require the removal of all confounding. We may be interested in whether we can transport treatment results to adopters of a new therapy but not all others continuing on the older therapy. Relationships between multiple unmeasured or incorrectly specified factors and the outcome could cancel one another out, resulting in matching outcomes for one treatment arm with remaining unmeasured effect measure modifiers. It is also possible for causes of treatment effect heterogeneity to be associated with

the outcome only in individuals receiving one of the two treatment arms. This approach is inherently retrospective and assesses whether trial results were transportable to a population that has already had a treatment decision made for it. This further solidifies its role as a method for ruling out transportability, as being able to transport to a claims-based population from 2010 to 2014 is no guarantee of transportability to a claims-based population from 2014-2018 while being unable to transport to that 2010-2014 population is likely troublesome for subsequent effect transport.

2.6: Public Health Significance

Understanding the risk-benefit profile of dabigatran in older adults is important for public health. Oral anticoagulants, including warfarin and dabigatran are used frequently in patients with AF, especially as criteria have shifted over time and point-of-care and self-INR monitoring have become more common.[98,99] Since their introduction to the market in 2010, novel oral anticoagulants (including dabigatran) account for a large share of oral anticoagulant prescriptions[100-103] and an even larger share of anticoagulant cost in the United States and worldwide, with dabigatran ranking 88th in current sales rank for all U.S. pharmaceuticals with over 800,000 prescriptions quarterly in 2013.[104] With a large patient population and such frequently prescribed drugs, it is vitally important to make sure that the benefit-risk distribution for the potential treatment options are being appropriately presented to stakeholders at the patient, provider, and insurer level.

This is even more important when discussing older adults participating in Medicare with higher risks of both adverse outcomes from AF (strokes) and adverse outcomes from anticoagulants (bleeding). Additionally, if there are important differences in outcomes between warfarin patients in trials and warfarin patients in routine care, even after accounting for differing patient characteristics with weighting methods, both trials and target population data repositories may need to become more flexible and record more data on potential effect modifiers going forward in order to be relevant for public health. Synthesizing trial and non-experimental data represents an important step forward for patients with AF and public health by allowing better understanding of the potential risks and benefits of treatments with warfarin and novel oral anticoagulants.

CHAPTER 3: RESEARCH PLAN

This research plan was designed to compare outcomes in older adults in routine care taking warfarin or dabigatran with risks in RE-LY trial participants, before and after adjusting for known risk factors; estimate effects of dabigatran compared to warfarin on ischemic stroke and gastrointestinal bleeding risk in older adults using transport weights; and assess how well estimates from various non-experimental methods agree with these transported trial results.

3.1: Overview

The aims called for estimating safety and effectiveness of starting dabigatran compared to starting warfarin in the U.S. Medicare population with atrial fibrillation (AF) through transporting the results of the RE-LY trial to several target populations. Because non-experimental approaches and direct use of the trial estimate both have their limitations, we fused cohort data from the Medicare population with the RE-LY trial data. Combining these two data sources made it possible to estimate average treatment effects in the Medicare population that take advantage of the randomized nature of RE-LY. This estimate would be valid under the assumption that a sufficient set of causes of sampling into the trial that are risk factors for the outcome have been accounted for.[105] The outcome data collected on patients in Medicare allowed us check this assumption using various weighting techniques as well as assess how different the results are from results obtained using non-experimental approaches.

3.2: Data Sources and Collection

This research used two data sources from two populations that were collected in very different ways: the RE-LY trial's individual-level data and cohorts of initiators of dabigatran and warfarin in Medicare from 2010 to October 2015.

3.2.1: RE-LY Population

3.2.1.1: RE-LY Population Description

Individual-level data for the RE-LY trial was obtained using Clinical Study Data Request (CSDR) and analyzed on the Clinical Trial Data Transparency (CTDT) platform, soliciting data access as soon as the analysis protocol is finalized. The RE-LY trial randomized 18,113 patients to warfarin, 110 mg dabigatran twice daily, and 150 mg of dabigatran twice daily. We focused primarily on the results for the 150 mg dosage of dabigatran, since the 110 mg dose was not approved for usage in the United States. Eligible patients had documented AF during the six months before their enrollment, as well as at least one of five other risk factors for stroke including: age greater than 75; previous stroke, TIA, or systemic embolism; left ventricular ejection fraction under 40% in the past six months; a diagnosis of diabetes mellitus and age over 65; or hypertension requiring pharmaceutical treatment and age over 65; or documented coronary artery disease and age over 65.

Key exclusions to assure either safety or efficacy of warfarin or dabigatran included reversible AF, prosthetic heart valves or other conditions for which dabigatran had not been tested, stroke within the past 14 days or severe stroke within the past 6 months, a variety of conditions associated with increased risk of bleeding, active infective endocarditis, active liver disease, anemia or thrombocytopenia, patients judged unreliable or having a life expectancy less than the expected trial duration, patients who received another investigational drug within 30 days, transaminase elevations in response to ximelagatran (another agent with a similar mechanism of action), and patients with severe renal impairment (creatinine clearance of equal to or less than 30 mL/min). The RE-LY trial was

conducted in a population from a variety of countries with a wide range of ages (mean age in dabigatran 150 mg of 71.5 years, standard deviation 8.8) and collected data on multiple potential causes of heterogeneity including medication use, past stroke, and other medical diagnoses.

As a randomized controlled trial there is no confounding in expectation in the baseline covariate distribution in RE-LY (and analyses suggested limited chance confounding). This allowed estimation of an internally valid intention-to-treat effect estimate in the target population using outcome data from RE-LY in Aim 1 provided we have weighted the RE-LY data to match the distributions of measured effect modifiers in our target. It does not, however, guarantee that an estimate censoring at treatment discontinuation will be internally valid. We limited our population to patients over 65 from the RE-LY trial to ensure we are looking only at older adults; this was not problematic from a sample size perspective, as 85% of the initial RE-LY trial population was over the age of 65.

3.2.1.2: Outcome Assessment in RE-LY

The RE-LY trial followed individuals for two primary outcomes after treatment initiation: first stroke or systemic embolism (efficacy) and first major hemorrhage (safety). These outcomes were reviewed and categorized by an international team of blinded adjudicators and patients were also provided symptom questionnaires at regular intervals; these symptom questionnaires were followed up on with medical record review. Stroke was defined as "sudden onset of a focal neurologic deficit in a location consistent with the territory of a major cerebral artery" and was divided into ischemic, hemorrhagic, and unspecified types. Major bleeding was defined as reduction in hemoglobin level of 20 grams per liter or more, transfusion with at least 2 units of blood, or symptomatic bleeding in a critical area or organ.

Within the context of our study, we used RE-LY's outcome data on first ischemic stroke and ignored the systemic embolism outcomes (as it is difficult to compare these results given difficulties

capturing embolisms in claims data)[106]; risk of hemorrhagic stroke may be elevated with a more potent anticoagulant while risk of an ischemic stroke is lowered, so all stroke were analyzed in a secondary analysis. When assessing bleeding outcomes, focused specifically on gastrointestinal bleeding events, with the less-specific major bleeding term being used for secondary analyses. We also used crude mortality data from the trial.

3.2.1.3: Exposure Assessment in RE-LY

In order to facilitate comparison with non-experimental analyses and prevent the large gap in warfarin adherence (see section **2.3.2: Non-experimental Studies** under Background and Significance) from jeopardizing transportability, our main analyses censored individuals at treatment switching or discontinuation.

In order to conduct any non-ITT analysis, we required data on when individuals discontinued or switched from drug to which they were randomized. Fortunately, the RE-LY trial tracked individuals to see whether they discontinued the study drug across the study period, with a one week grace period after permanent discontinuation. In our intention-to-treat secondary analyses, exposure assessment for trial patients was trivial: all their person-time was assigned the exposure to which they were randomized.

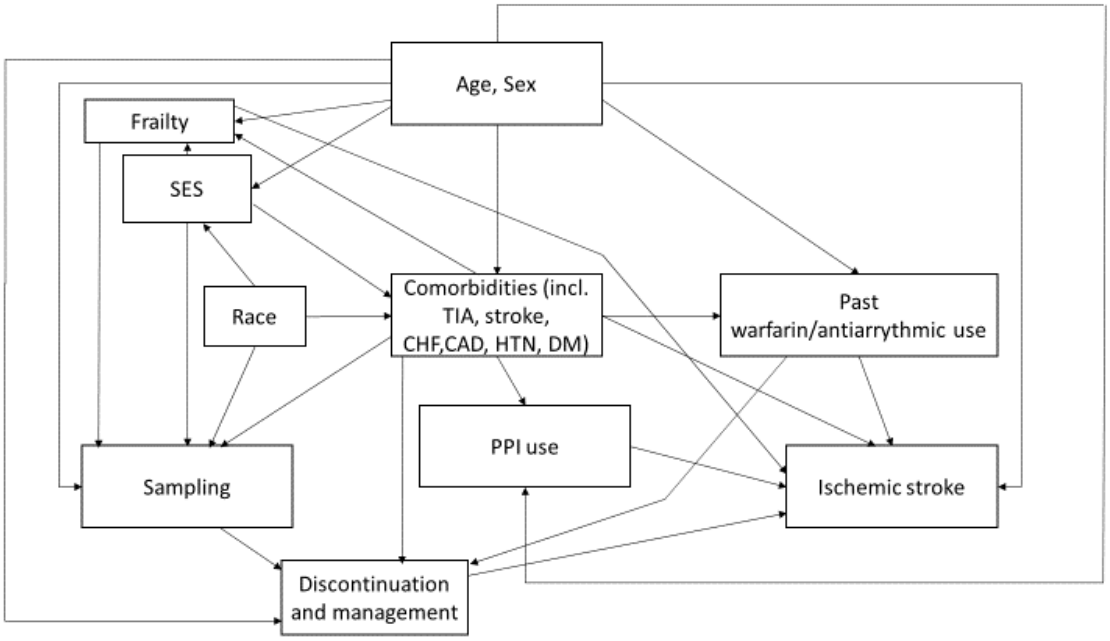

Figure 8: Causal diagram showing the association between sampling, ischemic stroke, and covariates.

3.2.1.4: Covariate Assessment in RE-LY

Two types of covariates were assessed in RE-LY in addition to exposure and outcome: 1) effect measure modifiers for use in the sampling model and 2) variables associated with discontinuation and censoring. Fortunately, RE-LY collected a large quantity of baseline information on variables hypothesized to be associated with the outcome in order to examine treatment effects in various subgroups that can be leveraged in these analyses.

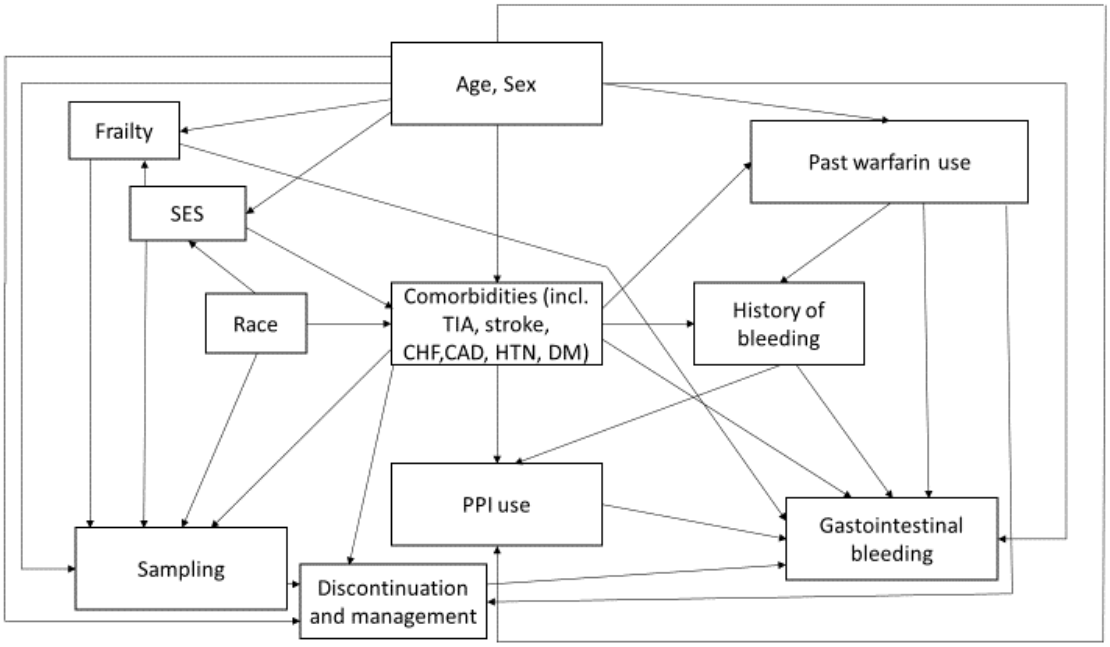

Figure 9: Causal diagram depicting the association between sampling, gastrointestinal bleeding, and other covariates.

Effect measure modifiers: Potential effect measure modifiers were identified from **Figures 8 and 9**, which depict hypothesized causal relationships between sampling, our outcomes of stroke and bleeding, and a variety of other variables. These diagrams were built by examining the marginal distributions of variables in trial and target populations (including in the literature review) as well as review of risk factors for stroke and bleeding for patients with AF. As mentioned in section **2.5.2: Weighting Methods**, a transport model that renders sampling independent of the outcome should allow estimation of an unbiased estimate in the target population. After adjusting for comorbidities, age, sex, frailty, and history of past medication use, the only path from sampling to either of our outcomes will be via the discontinuation and management node. Fortunately, censoring individuals at discontinuation or switching will reduce the potential for differing rates of discontinuation to be problematic. Unfortunately, without detailed lab and clinic data in both the trial and target population removing the potential for differences in warfarin management is unlikely to be possible. Additionally, several

variables had large differences between RE-LY and Medicare, likely due to a combination of differing sensitivity and specific of our pre-specified code-based algorithms compared to clinical assessment with longer lookback (when examining past vitamin K antagonist use) or simply assessing different things (when looking at bleeding in the past year vs. bleeding while on a vitamin K antagonist). This made weighting by hypertension, history of bleeding, and past use of vitamin K antagonists generate extremely large weights that inflated variance unreasonably. Our main analyses omitted these variables, though we did explore using an altered definition of hypertension. Still, **Aim 2.1** assessed the extent to which this or other unmeasured factors may bias our prediction of the outcome in the target population.

Censoring weight covariates: We used baseline covariates to standardize the population continuing treatment to continue to be representative of the population at baseline even if some of the effect modifying covariates are associated with censoring (or, in this case, trial discontinuation). This is for two reasons: first, there is potential for introduction of confounding after initial randomization. If, for example, patients with hypertension are more likely to discontinue treatment with warfarin than treatment with dabigatran and hypertension is associated with increased stroke risk, an unweighted effect estimate censoring at discontinuation or switching would be biased in favor of dabigatran.[107] Inverse probability of discontinuation weights can help deal with this problem.

The second reason is particular to transporting effect estimates that requires continuous prescribing and deals with mismatched effect measure modifiers or risk factors in the two populations. Suppose that patients in the target population with diabetes in the target population are more likely to discontinue their warfarin or dabigatran, while patients in the trial population are not. If we do not weight or standardize both populations, we could see differences in our weighted trial effect estimate and non-experimental population effect estimates purely because at later time points the trial population includes more individuals with the effect measure modifier of diabetes. We could also see

differences in our **Aim 2.1** assessment of transportability that arise only because of differences we've induced by requiring individuals to stay on their initial therapy.

The solution was stabilized inverse probability of censoring (sIPCW) weights. In the RE-LY trial, these were estimated with stepwise logistic regression within each treatment arm assessing the probability of staying on treatment up to a given time point, then the next time point, then the next, and so on. Variables were assessed at baseline, as RE-LY did not have readily available data on post-baseline variables. Individuals time periods were then assigned weights according to the following equation where Z_i is the set of time-varying and baseline covariates associated with sampling and Tx is the treatment arm.

$$sIPCW = \frac{P(Censored_t = 0 |\ Censored_{t-1} = 0, Tx = X\ )}{P(Censored_{t=0}\ |\ Censored_{t-1} = 0,\ Z_i, Tx = X)}$$

We included effect measure modifiers as well as confounders that are known in the trial (discussed below in the non-experimental section) in the model. Fortunately, warfarin and dabigatran do not cause changes in many of these variables, particularly most of the comorbidities and age, so there are few mediators we have to worry about conditioning on.

3.2.2: Medicare Atrial Fibrillation Population

3.2.2.1: Medicare Atrial Fibrillation Population Description

The specific Medicare data used was 20% sample of all Medicare beneficiaries with fee-for-service coverage of Medicare Parts A, B, and D for at least one month from 2008-2015, available at the University of North Carolina through the Sheps Center. We constructed two main study cohorts from the Medicare population from 2010 (when dabigatran was approved for use in the United States) to October 2015 (when the United States transitioned from ICD-9 to ICD-10 codes) for this analysis: first, a

cohort of dabigatran and warfarin initiators; and second, a cohort of dabigatran and warfarin initiators, where dabigatran users are allowed to have previously initiated warfarin or switched to dabigatran.

Patients were eligible for inclusion into the Medicare cohort at their first initiation of warfarin or 150 mg twice daily dabigatran with a 60-day washout period for use of either drug or another NOAC (apixaban, rivaroxaban, and edoxaban), provided they had an inpatient or outpatient AF diagnosis code in the 6 months before or 1 week after their prescription (in which case follow-up began at the time of diagnosis). This requirement for a recent AF diagnosis code was analogous to the RE-LY trial inclusion criteria that required evidence of recent AF. Individuals had to meet the eligibility criteria for RE-LY including at least one risk factor for stroke (described in section **3.2.1.1: RE-LY Population Description**), as represented by at least one diagnosis code in the year prior to initiation (specific codes listed in **Appendix B**) or age over 75. Individuals also needed to have 12 months continuous coverage in Medicare parts A, B, and D before this index prescription to enable assessment of eligibility criteria, key effect measure modifiers, and potential confounders.

This period was also used to exclude individuals with identifiable exclusion criteria for the trial, including liver disease, severe stroke within the past six months, anemia and thrombocytopenia, valvular AF or prosthetic heart valves, and severe renal insufficiency (specific codes listed in **Appendix B**). If there were multiple eligible initiations, only the first eligible initiation was included. This Medicare cohort contributed external validity to the project and allowed insight into the distribution of effect modifiers in a general clinical cohort participating in anticoagulation care that may not have been willing or able to participate in a trial, critical for **Aim 2**. It also provided data for the **Aim 1** non-experimental analyses.

The Medicare cohort was further divided into four potential target populations: 1) patients initiating warfarin for AF that could have been included in RE-LY; 2) patients initiating dabigatran for AF that could have been included in RE-LY; 3) patients with less than 15% predicted probability of frailty

that initiated warfarin for AF that could have been included in RE-LY; and 4) patients with less than 15% predicted probability of frailty that initiated dabigatran for AF that could have been included in RE-LY.

3.2.2.2: Outcome Assessment in Medicare

Unfortunately, we did not have access to medical records for review by an international blinded group of adjudicators for use in this cohort. Instead, we used ICD-9 codes in claims, both inpatient and outpatient, to identify ischemic stroke and gastrointestinal bleeding. We used similar ICD-9 codes to identify the types of stroke that were used in past non-experimental analyses using claims-based data to facilitate comparisons to their results (i.e., by Seeger et al.).[18] Medical record review in some databases have shown positive predictive value of close to 90% for these codes, suggesting they perform quite well.[108] The full list of codes for stroke is presented in **Appendix A.** We specifically examined codes for ischemic stroke, with secondary analyses examining all strokes.

When identifying major bleeding events, we a[illegible] ICD-9 codes (presented in the second portion of **Appendix A**) in inpatient and outpatient claims. These definitions were similar to and adapted from the definitions used in the other non-experimental studies in claims-based data, particular Seeger et al, and map directly to the trial outcomes.[18] These codes and definitions have been shown to have positive predicted values between 80% and 90% with medical chart review in claims databases, particularly in the setting of anticoagulant-associated adverse events.[109,110] Than main analysis was carried out separating gastrointestinal hemorrhage from other major bleeding, with all major bleeding examining secondarily.

3.2.2.3: Exposure Assessment in Medicare

Prescription claims data for Medicare beneficiaries were used to identify warfarin and dabigatran initiators from 2010 to October 2015. Individuals were defined as initiators if the days supply from their last prescription for an oral anticoagulant ran out at least 60 days prior to the initial prescription. In the analysis censoring at discontinuation individuals' follow-up time was censored after switching medications as or having a 30-day gap in novel oral anticoagulant coverage or a 45 day gap in warfarin coverage as shown in **Figure 10**, with the larger gap for warfarin provided due to the fact that pharmacy days' supply may be out of sync with the way patients are taking warfarin due to changes in directions for use at anticoagulation management appointments. To help identify warfarin initiators who may start to pay purely out of pocket, CPT codes for INR draws or anticoagulation management (CPTs 85610, 99363, and 99364)[111] "refreshed" warfarin prescriptions and extended the length of follow-up for thirty days from the time of the CPT code. In the intention-to-treat analysis, individuals will be followed until death.

Figure 10

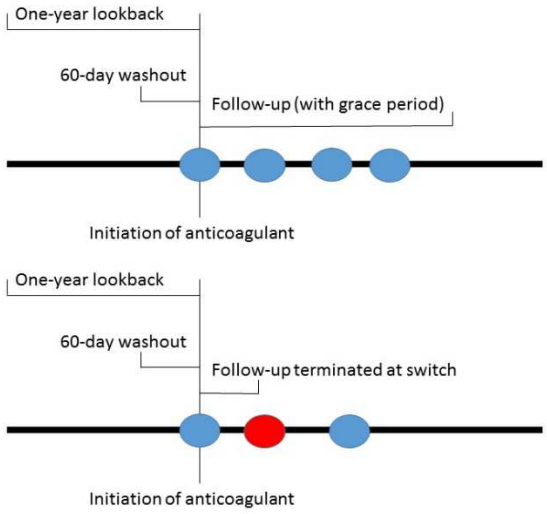

Figure 10: Follow-up and lookback structure for the primary analysis.

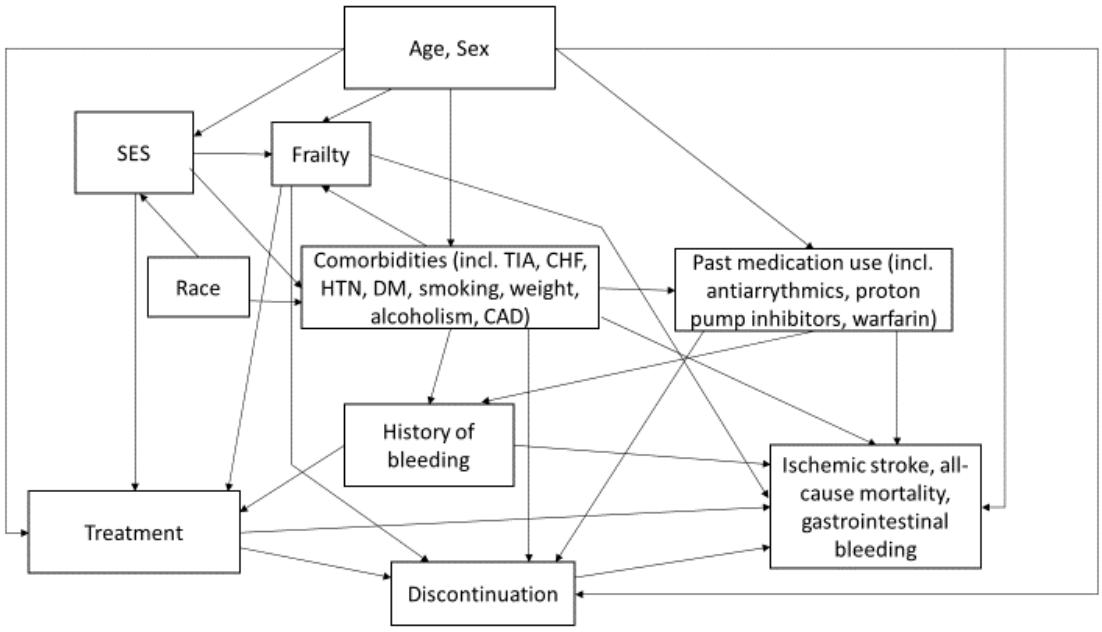

Figure 11: Directed Acyclic Graph (DAG) showing causal relationships between treatment, covariates, and outcomes.

3.2.2.4: Covariate Assessment in Medicare

We assessed three kinds of covariates in the Medicare cohort: 1) effect measure modifiers, 2) confounders, and 3) censoring covariates.

Effect measure modifiers: Effect measure modifiers in the Medicare cohort were assessed at the date of treatment initiation for the Medicare cohort. We used a lookback period of one year in claims data to assess the presence of the variables depicted in the sampling diagrams depicted in **Figure 8** and **Figure 9**. Specific ICD-9 codes are listed in **Appendix B.**

Confounding variables: Confounding variables were assessed at the date of treatment initiation for the Medicare cohort with the same one-year and all-available lookback approaches as the effect measure modifiers. The specific variables included were based upon a directed acyclic graph, **Figure 11**, which indicates which variables form a minimally sufficient adjustment set in the Medicare cohort that

close all open backdoor paths between treatment and the outcomes. Because the set of variables that affect treatment choice are identical with respect to both outcomes and ischemic stroke and gastroinestinal bleeding events share many risk factors, the same adjustment set was used in each analysis. The set of variables can be divided into three main categories: demographics (age, sex, race, and socioeconomic status), comorbid conditions (transient ischemic attack, congestive heart failure, diabetes, smoking, weight, alcoholism, history of bleeding, and frailty), and past medication use (past warfarin use). Definitions and codes used to identify each of these is presented in **Appendix B**.

The associations between the demographics, comorbidities, and outcome was built based upon literature review of the various epidemiological studies in section **2.3.2: Non-experimental Studies** and cardiovascular and risk scores including the $CHADS_2$ score[112] and the Framingham Risk Score.[113] The associations between each of these variables and whether a patient might use dabigatran rather than warfarin came from questions regarding prescribing preferences to a medical professional working in an anticoagulation clinic, section **2.3.2: Non-experimental Studies**, and treatment guidelines.[5] As can be seen in the graph, analyses requiring individuals to stay on treatment in this context involve removing a potential mediator of treatment effect (potential for discontinuing the treatment in question).

Censoring weight covariates: We also built censoring weights using effect measure modifiers and confounders in the Medicare population for use in analyses conditional on staying on initial treatment and ensure that the population that continues on treatment is standardized to look like the initial population in the Medicare target population; otherwise, as discussed in section **3.2.1.4: Covariate Assessment in the RE-LY Trial,** selective dropout in this estimate could lead to either bias in internal validity (only healthy individuals stay on warfarin, while unhealthy users stay on both medication) or external validity (the target and trial populations change differently over time from dropout). Weights were estimated separately within the Medicare population from the RE-LY population

because the processes leading to discontinuation are likely quite different between the two populations; the discontinuation rates are much higher in claims than RE-LY.

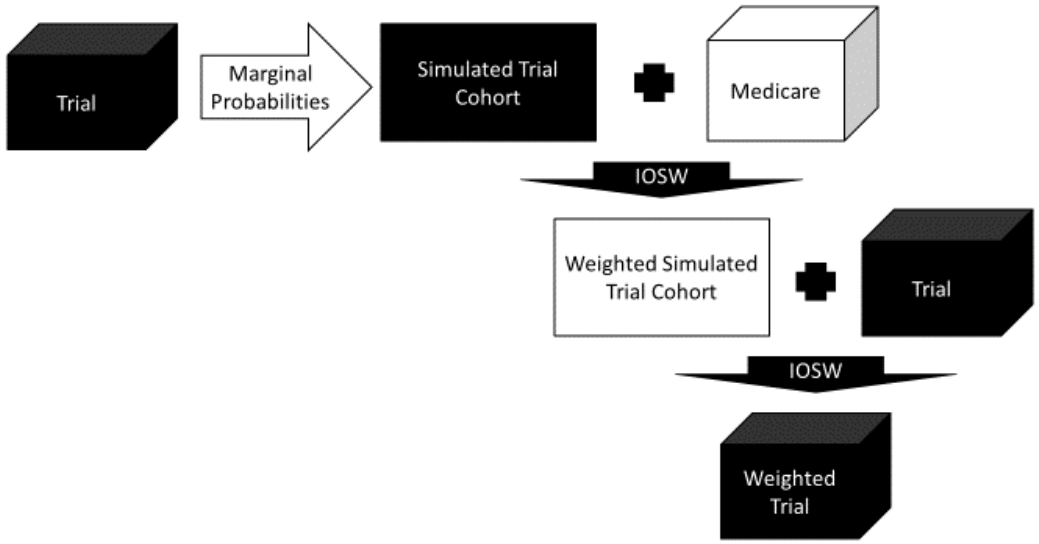

Figure 12: Data access plan for combining data from the two populations in Aim 1 and 3.

3.2.3: Data Combination

Unfortunately, we were not able to combine our two data sets at an individual level (the ideal gold standard) for analysis because of data ownership and privacy concerns. Medicare does not allow data to leave their server in tables with cell size greater than 11 and is reluctant to allow individuals to pull data sets with the exact timing of outcome and censoring events. On the other hand, few drug companies are willing to have individual-level patient data leave their system that isn't being monitored, and our original plan to only estimate distributions of effect modifiers did not satisfy their restrictions on individual-level data. At first glance, this rendered it difficult to conduct **Aim 1** without making enormous and potentially incorrect assumptions about correlation between variables, but we used the data plan in **Figure 12** as a substitute for directly combining data between the populations.

First, when estimating our sampling weights and transporting estimates, no data on effect modifiers left the Medicare platform. Instead, we took marginal distributions of variouse modifiers and

the joint distribution of age and sex and used these to create a simulated cohort of 300,000 individuals, which was weighted to resemble the various Medicare target populations using main effects and one-way interactions for all variables, modeling age with quadratic restricted splines with four knots at the 20^{th}, 40^{th}, 60^{th}, and 80^{th} percentile. Age was also capped at 90 due to the way trial handled these extreme ages. The coefficients from this model were used on a recreated simulant population on the trial server, which was in turn used for the target population. While these methods may create issues when there are three-way interactions between modifiers, the result did balance marginal distributions of modifiers quite well. Additionally, we did not directly export any survival curve data from the Medicare server. Instead, we exported the percent of individuals alive at the end of each week and at set bench marks (180 days, 365 days, 545 days, and 730 days) for comparison with the trial data in **Aim 2.1.** Bootstrapping took all this into account with the modifiers, sampling models, and outcomes all being estimated with the same seeds.

3.3: Data Analysis

We 1) estimated two-year risk differences using non-experimental epidemiologic methods from only the Medicare AF population and 2) estimated two-year risk differences comparing warfarin from NOACs in the Medicare AF population with transport methods and assessed transportability.

3.3.1: Aim 1, Estimating Non-Experimental Treatment Effects in Medicare (Figure 13)

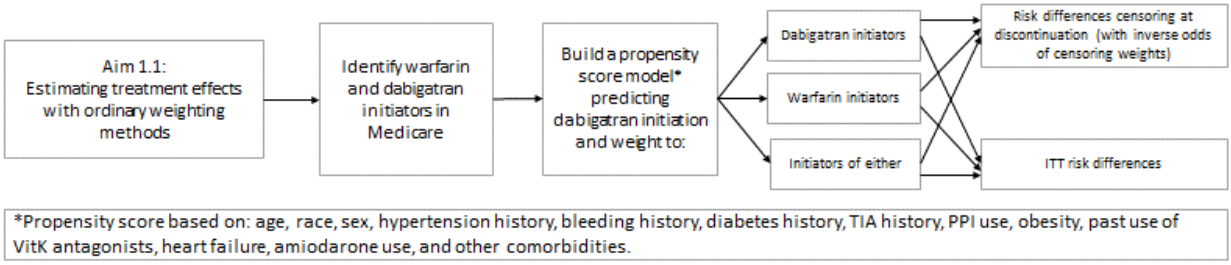

Figure 13: Aim 1 conceptual diagram.

Objectives: **Aim 1** focused on estimating the treatment effect on the two-year risk of ischemic stroke, death, and gastrointestinal bleeding for dabigatran versus warfarin in various Medicare target populations using multiple types of non-experimental methods.

3.3.1.1: Aim 1.1, New User Active Comparator Design

Objectives: In **Aim 1.1** we estimated risk differences of two-year ischemic stroke, death, and gastrointestinal bleeding for patients in the cohorts of dabigatran and warfarin initiators using a new user active comparator design in the Medicare cohorts to be used as target populations in **Aim 2.**

Methods (inverse probability of treatment weights): We balanced the confounding variables between dabigatran and warfarin initiators using inverse probability of treatment weighting. To use these weights, we estimated the probability of dabigatran initiation based upon the confounders described in **Figure 11** fitting main effects and additional terms as required to achieve marginal[72] differences as small as possible, with any major confounders with an SMD greater than 0.100 being unacceptable (though this is a fairly arbitrary cut point and none came close to this threshold in the initial model).[114] Stabilized inverse probability of treatment weights (sIPTW) were assigned based upon

the model predicted probabilities.[115] Dabigatran initiators were weighted using equation 2 and warfarin initiators with equation 3, where Z_i represents the confounding variables used in the propensity score estimation.

$$\textit{Equation 2: } sIPTW_{Dabi} = \frac{P(dabigatran)}{P(dabigatran \mid Zi)}$$

$$\textit{Equation 3: } sIPTW_{Warf} = \frac{1-P(dabigatran)}{1-P(dabigatran \mid Zi)}$$

Methods (other statistical considerations): The outcomes of interest were compared by contrasting the IPTW and IPCW-weighted survival curves (constructed with Aalen-Johansen methods and the above weights for each individual) in the Medicare population at one and two years after applying sIPCW. We bootstraped 200 replicates to estimate confidence intervals. Since sIPTW estimated the treatment effect in the entire population and we desired treatment effect estimates in the dabigatran and warfarin users for comparison with the estimates obtained with sIOSW, we will also use standardized mortality ratio (SMR)[116] weights to estimate treatment effects in dabigatran and warfarin users specifically.

3.3.2: Aim 2, Transporting Treatment Effects with Inverse Odds of Sampling Weights (Figure 14)

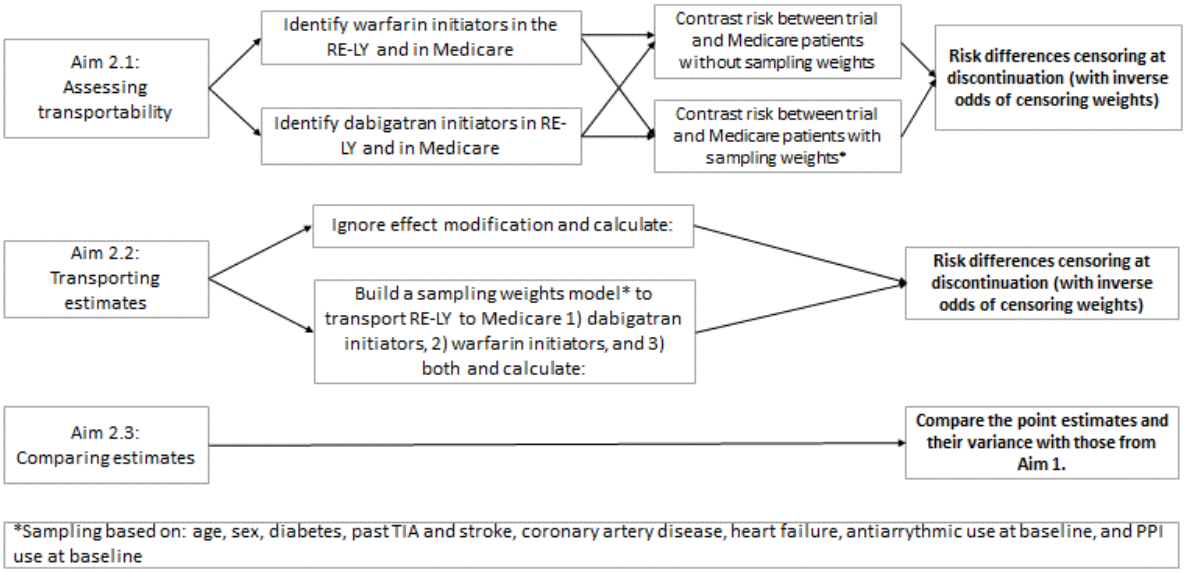

Figure 14: Aim 2 conceptual diagram.

Objectives: **Aim 2** was focused on first assessing whether we might be able to transport treatment effects and then actually estimating treatment effects on the two-year risk of ischemic stroke, all-cause mortality, and gastrointestinal bleeding for dabigatran versus warfarin in the Medicare target population using sIOSW.

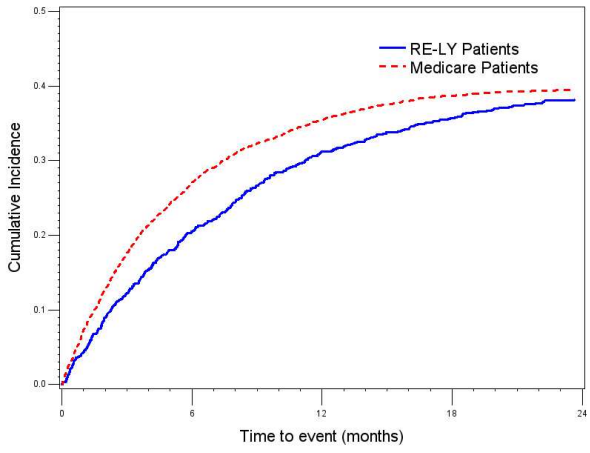

Figure 15: Hypothetical graph comparing cumulative incidence of bleeding for RE-LY and Medicare patients. The groups diverge in cumulative incidence quickly.

3.3.2.1: Aim 2.1, Assessing Transportability

Objectives: For **Aim 2.1**, we used weighting methods to assess the degree to which the assumptions necessary for transport of treatment effects on the two-year risk of ischemic stroke, death, and gastrointestinal bleeding for dabigatran versus warfarin have been met in two target

populations with methods from indirect trial comparisons and control arm assessment.[96,117]

Methods (initial analysis): The first step in this process was comparison of unadjusted risk of stroke, bleeding, and death in the individuals randomized to dabigatran in RE-LY to the unadjusted risk of each of the outcomes in the Medicare dabigatran patients (with a similar comparison for the warfarin patients in the RE-LY trial and Medicare). This comparison was conducted both visually by overlaying the unadjusted Aalen-Johansen cumulative incidence curves (see **Figure 16**) as well as quantitatively by computing risk differences at 6, 12, 18, and 24 months. Variances were obtained with bootstrapping with 200 replicates (fewer than the 1000 originally planned due to the time taken by the competing risks macro), bootstrapping populations before limiting to complete cases or performing subsample validation. For both groups, follow-up was censored at discontinuation or switching and inverse probability of selection weights were used to correct for potential bias.

Methods (inverse odds of sampling weights): We also assessed whether weighting by measured effect measure modifiers reduced the disparity in outcomes between the groups,. We constructed weights from a model assessing sampling into the RE-LY trial from the Medicare AF cohort within strata of exposure X (where X is either warfarin or dabigatran) in the target population; since exposure was randomized within the trial we can treat the trial cohort as having both exposures when building this model. Odds weights were necessary so that the trial population is not assumed to be part of the target population (see section **2.5.2: Weighting Methods**). If individuals were on treatment X in the RE-LY trial, they are weighted according to equation 4 with the probabilities being calculated from a logistic regression model, where X is either warfarin or dabigatran and EMM represents the set of all selected effect measure modifying covariates and interaction terms.

$$sIOSW = \left(\frac{P(RE-LY\ trial \mid X)}{1-P(RE-LY\ trial \mid X)}\right) * \left(\frac{1-P(RE-LY\ trial \mid EMM, X)}{P(RE-LY\ trial \mid EMM, X)}\right)$$

Individuals in Medicare received weights of 1. We used quadratic restricted splines for age when fitting the logistic regression models to account for complex types of effect measure modification.

Other statistical considerations: We repeated the survival analysis using these weights to construct the Aalen-Johansen cumulative incidence curves for the RE-LY trial and Medicare cohorts (separately for each anticoagulant) and compared them visually and quantitatively with risk differences at 6, 12, 18, and 24 months just as we did the crude cumulative incidence curves. Variances were obtained with bootstraps of 200 replicates, making sure to bootstrap both populations before taking complete cases or conducting any form of imputation or subsample validation.

If the weighted trial and target population curves differ and there is a non-null treatment effect and the outcomes are measured identically, it is impossible to have taken into account all modifiers that differ between the groups on both scales, making it less likely all modifiers of the risk difference have been controlled for and increasing the possibility that the assumptions necessary for transportability have been violated. These analyses were particularly important for evaluating the performance of warfarin initiators in Medicare relative to warfarin patients in the RE-LY trial. We also assessed transportability after limiting ourselves to patients in the Medicare cohort with under 15% predicted probability of frailty based on the frailty prediction algorithm created by Faurot et al.[118] to help eliminate potential for including individuals in our target populations that were not trial-eligible due to high frailty or short life expectancy. Since these frailty-restricted Medicare populations have more similar outcomes to the RE-LY trial participants than the original Medicare cohort, they were used for additional target populations in **Aim 2** and also re-assessed in **Aim 1.**

3.3.2.2: Aim 2.2, Transporting Treatment Effects

Objectives: In **Aim 2.2**, we estimated treatment effects on two-year risk of ischemic stroke, death, and gastrointestinal bleeding for dabigatran versus warfarin in various Medicare target populations using stabilized inverse odds of sampling weights (sIOSW) and data from the RE-LY trial.

Weighting methods: Normal inverse probability of treatment weights methods typically eliminate confounding by creating exposed and unexposed pseudo-populations with identical confounder distributions after estimating the probability of exposure condition on various covariates. With IOSW, we instead transformed the RE-LY trial population into a pseudo-population with an identical distribution of baseline effect measure modifiers as the Medicare AF population[87] by combining the joint categorical modifier RE-LY data and Medicare AF data and using logistic regression to identify the probability of selection based upon the selected effect measure modifiers. If individuals were enrolled in RE-LY, they received stabilized inverse odds of sampling weights (sIOSW) based on this equation:

$$(sIOSW) = \left(\frac{P(RE-LY\ trial)}{1-P(RE-LY\ trial)}\right) * \left(\frac{1-P(RE-LY\ trial \mid EMM)}{P(RE-LY\ trial \mid EMM)}\right)$$

Unlike in **Aim 2.1**, Medicare patients received weights of 0 instead of 1 and the estimates were no longer conditional on the specific type of treatment received; X has dropped out of the equation. When weights became large when including several variables that may have been due to discrepancy in measurement or actual differences in distribution (hypertension and vitamin K antagonists), we omitted them. In the end we examined the marginal covariate balance using standardized mean differences between the target and RE-LY population.

Follow-up and censoring: To reduce potential issues with differential persistence between the trial and target population estimates influencing the comparison, we focused on an analysis censoring trial individuals at treatment switching or discontinuation. Stepwise inverse probability of censoring weights based upon the set of EMM and confounders as described in **3.2.1.4: Covariate Assessment in**

the RE-LY Trial and fit separately in the trial and target populations for each treatment were used to prevent potential selection bias in this approach. We also explored an intention-to-treat design as a secondary analysis.

Other statistical considerations: The outcomes of interest were compared on the risk difference scale by directly comparing the weighted survival curves in the RE-LY trial population (constructed with the above weights and Aalen-Johansen methods rather than assuming we can prevent competing risks[119]). The weighting process was bootstrapped with 200 iterations in order to generate 95% confidence limits for the risk differences, with both the trial and target populations being bootstrapped before any imputation, limitation to complete cases, or subsample validation. This process was repeated for each of the target populations of interest (warfarin initiators, dabigatran initiators, and those with lower predicted probability of frailty). When these estimates from **Aim 2.2** differed from the results of the RE-LY trial, it lends credence to the idea that there are meaningful effect measure modifiers in place that were differentially selected for by the trial enrollment process.

3.3.2.3: Aim 2.3, Contrasting Effect Estimates

Objectives: In **Aim 2.3** we compared treatment effect estimates for two-year risk of ischemic stroke, death, and gastrointestinal bleeding in various Medicare populations obtained using the data from the RE-LY trial and the new user active comparator design.

Methods: The estimates obtained under these **Aim 1** were directly contrasted with the estimates obtained under **Aim 2**. This process was repeated for each of the four target populations: 1) new users of dabigatran; 2) new initiators of warfarin; 3) less frail new users of dabigatran; and 4) less frail new users of warfarin. Comparisons were performed by plotting the point estimates and 95% confidence limits of the trial and non-experimental estimates against one another, with particular attention paid to shifts in estimate across the null between methods. If the results in **Aim 2** differed

from those in **Aim 1**, either additional or incorrectly modeled confounding factors exist (e.g. socioeconomic status has an unblocked path to the outcome), additional or incorrectly modeled effect measure modifiers exist (e.g. better management in the trial than the target population), different outcomes are being assessed in both populations, or all three.

3.4: Human Subjects

This research was conducted in two large databases and reviewed by the UNC institutional review board. We anticipated minimal potential harms to individuals in either of the two databases, with the only risk being potential infringement of privacy or identification. To ensure appropriate precautions are taken for preservation of patient privacy, data was kept on secure servers at all times. When the marginal distribution data was exported from the CTDT platform, it was downloaded directly onto the Sheps server. It was never downloaded onto a private computer. Similarly, the weekly cumulative incidence from the Medicare cohort was uploaded into the CTDT platform directly from the Sheps server, rather than downloaded onto a private computer and transferred. This use of secure servers minimized the potential for privacy harms to individuals enrolled in the trial or captured in the claims database. Additionally, the survival data and joint distribution data will be deleted from the Medicare and CTDT platform, respectively, as soon as final analyses are complete to preserve custody of the data.

CHAPTER 4: STAYING ON TREATMENT MATTERS: ESTIMATING EFFECTS OF WARFARIN VS DABIGATRAN IN MEDICARE

4.1: Introduction

Atrial fibrillation affects an estimated 33 million adults worldwide.[1] Even if individuals with atrial fibrillation are asymptomatic, atrial fibrillation strongly increases stroke incidence and resulting strokes are more frequently associated with death, hospitalization, and long-term disability than strokes in adults without atrial fibrillation.[2,3] Warfarin, the historical standard of care for stroke prevention in atrial fibrillation, is difficult to manage therapeutically due to its lengthy half-life and narrow therapeutic range.[120] To make matters worse, mismanagement can result in catastrophic bleeding events.[4] Novel oral anticoagulants are comparatively easy to manage (with the necessary adjustments being dosage changes in the presence of renal insufficiency) and have been approved after several trials demonstrated non-inferiority to warfarin administered with suggested management protocols.[5,44,45,47] One of the first novel anticoagulants to be approved in the United States, dabigatran, appeared more effective than warfarin at ischemic stroke prevention (HR 0.76, 95% C.I. 0.60-0.98) in the RE-LY trial, though it did cause an increase in gastrointestinal bleeding (HR 1.50, 95% C.I. 1.19-1.89). A small relative improvement in all-cause mortality was also observed (HR 0.88, 95% C.I. 0.77-1.00), though it was not a primary outcome; this magnitude of protective effect in a population with higher mortality than the trial population could result in substantial mortality benefit.

However, estimates of efficacy in these clinical trials are likely imperfect estimates of effectiveness in clinical care.[7,8] Patients selected into trials tend to be younger with fewer comorbidities than patients in the general population. These differences could modify the population average treatment effect.[9,10] To address concerns about this potential treatment effect modification in wider

populations, studies have used claims data to directly estimate the safety of novel oral anticoagulants compared to warfarin in clinical care and observed attenuated efficacy and differing safety profiles.[11-20] While these studies were generally well conducted, gaps in knowledge remain. Most relied on propensity score matching of warfarin users to patients taking dabigatran resulting in imperfect understanding of treatment effects in the entire population. Most estimated treatment effects on the relative scale using hazard ratios rather than the absolute scale, which can be misleading when it comes to assesing the overall benefit-harm balance.[121,122] Finally, most studies censored individuals at treatment discontinuation or switching without assessing whether this was differential with respect to treatment; they also refrained from estimating treatment effects taking into account the natural switching and discontinuation patterns of real world populations.

In this study, we aim to estimate absolute-scale treatment effects of dabigatran versus warfarin on ischemic stroke, all-cause mortality, and gastrointestinal bleeding considering various patient populations and adherence scenarios of interest using non-experimental data.

4.2: Methods

4.2.1: Study Population

This study was performed in the 20% sample of Medicare beneficiaries managed by the Centers for Medicare and Medicaid Services from 2010 to 2015. Individuals were eligible for the analysis after 365 days of continuous enrollment in Medicare A, B, and D. All participants were required to be over age 65 with at least one of the following risk factors for ischemic stroke: hypertension, diabetes, congestive heart failure, and past stroke or transient ischemic attack (defined by ICD-9-CM diagnosis codes in any position in the past year), and age over 75. Individuals with diagnosis codes indicating prosthetic heart valves, endocarditis, primary diagnoses indicating cancer in the past 180 days (as the RE-LY trial excluded those with active cancer in the past six months), active liver disease in the past year, or chronic kidney disease were excluded. Codes are listed in **Appendix B.** These inclusion and exclusion criteria parallel those used in the RE-LY trial to the extent possible.

4.2.2: Exposure

We used an active comparator new user study design. We defined "new use" as having no days' supply of any oral anticoagulant used for atrial fibrillation (warfarin, dabigatran, and the other novel oral anticoagulants on the market, apixaban, rivaroxaban, and edoxaban) in the 60 days prior to receipt of a warfarin or dabigatran prescription during the study period. To ensure we were examining the dosage of dabigatran studied in the RE-LY trial, we limited analyses to the 150 mg dosage in assessing both treatment initiation and continuation. Each initiation was analyzed separately for eligibility criteria and included only if there was at least one diagnosis code for non-valvular atrial fibrillation present in the 180 days prior to or 7 days after new use. Only individuals' first eligible drug initiation of either drug was included in the analysis.

After identifying new users, we conducted two types of analyses: an analysis in which individuals were followed under their initial treatment for each of the outcomes until the end of the study period or the end of their Medicare A, B, and D coverage, regardless of whether they continued use of their oral anticoagulant (first treatment carried forward analysis, FTCF, analysis); and an analysis where individuals were censored after a gap in therapy more than 30 days (dabigatran arm) or 45 days (warfarin arm) (adherence adjusted, AA, analysis). Procedure codes for anticoagulation management (listed in **Appendix D**) were used to extend coverage in the warfarin arm for 30 days in the event that warfarin claims were unobservable.[111] The longer gap in the warfarin arm was to accommodate potential dosage changes during warfarin management. Medication stockpiling was not allowed in our analysis as it would lead to inaccurate estimates of days' supply for warfarin users on multiple strengths of the drug.

4.2.3: Outcomes

This study examined three main outcomes: ischemic stroke, defined by previously validated[108,109] ICD-9-CM codes listed in Appendix B; death, defined by the Medicare date of death; and gastrointestinal bleeding, defined by previously validated[109] ICD-9-CM codes listed in **Appendix A**. We also examined all strokes and major bleeds to compare results with those of the trial and several of the past studies. Codes for all outcomes had to appear in the primary position of an inpatient encounter to exclude codes for follow-up or medical history of the event, and individuals with day 0 outcomes were excluded from that analysis to avoid the possibility of the outcome preceding the initiation. Follow-up was outcome-specific; that is, if someone experienced an ischemic stroke, then spent another three months on the drug, those three months would be included in the death, bleed, and gastrointestinal bleeding outcomes.

4.2.4: Covariates

In addition to the inclusion and exclusion criteria, we measured a large number of baseline covariates (also listed in **Appendix B)** in this analysis[18] using one-year lookback period from the index date and estimated the predicted probability of frailty using a Medicare claims-based algorithm developed by Faurot et al.[118] We constructed directed acyclic graphs[123] (**Figure 11**) for the outcomes using expert opinion and a review of the literature. From these graphs, our measured covariates (provided they are measured without error) form a sufficient set for estimation of an unbiased effect of treatment on the outcome. Age and sex were available for all individuals; since our other covariates were defined by the absence of insurance claims, there was no missing data per se.

4.2.5: Statistical Analyses

We estimated the predicted probability of dabigatran initiation among new users of dabigatran or warfarin conditional on the confounder sets from logistic regression, modeling age and frailty with restricted cubic splines with four knots at the 20th, 40th, 60th, and 80th percentiles. The resulting probabilities were used to construct inverse probability of treatment weighted (IPTW) dabigatran and warfarin cohorts (to estimate treatment effects in the whole population) as well as a standardized mortality ratio weighted (SMR weighted) version of the warfarin cohort (to estimate treatment effects in the dabigatran initiators).[124] We checked to ensure these weights properly balanced covariates by assessing whether marginal absolute standardized mean differences between the groups after weighting were less than 0.10.[114] In adherence adjusted analyses, we implemented inverse probability of censoring weights as the inverse probability of not switching or discontinuing (calculating these probabilities separately) based on time-varying versions of confounders at the 25th, 50th, and 75th percentiles of the censoring distribution.[125] Censoring due to end of Medicare A, B, or D enrollment or

the end of the study period was treated as random censoring, with no weights constructed for these censoring mechanisms.

After applying weights, we estimated the risk of each outcome via a weighted Aalen-Johansen estimator to take into account the competing risk of death[68] at one- and two- years for each treatment group, using the standard deviations from 200 replicate bootstraps[126] to obtain limits for the 95% confidence intervals. All statistical analyses were conducted using SAS 9.4 for Windows (Cary, NC, USA).

4.2.6: Sensitivity Analyses

We conducted a variety of sensitivity analyses. First, we varied the allowable gap between prescriptions within a treatment episode to be 7 days or 60 days for both treatment arms. Second, we ignored procedure codes rather than using them to refresh warfarin coverage. Third, we excluded individuals with any code for stroke in the primary position of an inpatient encounter in the past 6 months to attempt to emulate RE-LY's exclusion of those with severe strokes in that time period. Finally, we excluded individuals with a predicted probability of frailty over 10% to examine treatment effects in a less-frail population. This cut-point was chosen to be more aggressive than that from the frailty score's initial validation[118,127] to remove those at higher risk of both one- and two-year mortality from the study population.

4.3: Results

Figure 16 provides a flow diagram detailing inclusion and exclusion criteria in the cohort. Of the 393,684 total new use periods for dabigatran and warfarin, 98,388 met inclusion and exclusion criteria. After restriction to the first eligible initiation per individual from 2010-2015, we had a final cohort of 10,717 dabigatran new users and 74,891 warfarin new users for analysis. The distribution of various covariates in these individuals are listed in **Table 2**. Compared to warfarin new users, dabigatran new users were generally younger (28.9% age 80 and over vs 43.3%) and more likely to be women (49.6% vs 43.3%), with lower predicted probability of frailty (median 0.050 vs 0.074), fewer codes indicating past bleeds (8.4% vs 12.8%), and fewer having used warfarin in the past (20.1% vs 35.7%). After IPTW or SMR weighting, baseline covariates were more balanced and absolute standardized mean differences (ASMDs) for each measured covariate were all less than 0.100 (see **Figure 17**). The mean of the stabilized weights was 1.002, with 15 of the older and frailer dabigatran new users having weights greater than 10 and one individual with a particularly high comorbidity burden having a large weight of 42.

Table 3 includes the rates and risks of ischemic stroke, all-cause mortality, and gastrointestinal bleeding across the various populations of interest and adherence scenario. Dabigatran users had shorter durations of time on treatment (median 152 days vs. median 259 days), with 59% of dabigatran users stopping treatment and 16% switching treatment during the study period compared to 44% of warfarin users stopping treatment and 8% switching treatment. Outcomes were common in the first two years of follow-up in the first treatment carried forward analysis: we observed unadjusted incidence rates of 1.37 (dabigatran) vs 1.66 (warfarin) per 100 person-years for ischemic strokes, 5.36 (dabigatran) vs 9.51 (warfarin) per 100 person-years for all-cause mortality, and 1.93 (dabigatran) vs 2.05 (warfarin) per 100 person-years for gastrointestinal bleeding events. The adherence adjusted analysis showed lower rates for ischemic stroke (0.91 (dabigatran) vs 1.56 (warfarin) per 100 person-years) and all-cause

mortality (3.59 (dabigatran) vs 8.08 (warfarin) per 100 person-years), but slightly elevated risks of gastrointestinal bleeding (2.27 (dabigatran) vs 2.27 (warfarin) per 100 person-years). **Figure 18** shows IPTW-weighted survival curves for ischemic stroke under FTFC and AA methodologies and illustrates the stark shift in adjusted absolute risk for dabigatran patients under the FTFC design compared to the AA; **Figures 19-22** contain survival curves for the other outcomes.

Table 4 depicts two-year risk ratios and risk differences under the AA and FTCF follow-up methods, both in the crude and after implementing the IPTW, SMR, and inverse probability of censoring weights in the AA. In AA analyses, dabigatran new use compared with warfarin new use was associated with fewer ischemic strokes (two-year RD: -0.73%, 95% CI -1.40%, -0.06%) and lower mortality (two-year RD: -2.98%, 95% CI -5.05%, -0.91%). There was an elevated risk of gastrointestinal bleeding, however (two-year RD: 1.79%, 95% CI -0.13%, 3.71%); the larger variance for this outcome is partly due to a bleeding event in the person with very large weight.

In IPTW FTCF analyses, the association between dabigatran new use and all-cause mortality was attenuated relative to the AA analyses (two-year RD: -0.84% -2.39%, 0.72%) and risk of ischemic stroke actually increased in dabigatran new users (two-year RD: 0.44%, 95% CI -0.22%, 1.09%). The increase in the risk of gastrointestinal bleeding was also attenuated compared to the AA analyses (two-year RD: 1.05%, 95% CI 0.08%, 2.01%).

SMR-weighted risk differences were similar to the IPTW ones, though they were attenuated for AA gastrointestinal bleeding (two-year AA RD: 0.51%, 95% CI -0.30%, 1.31%) and farther from the null for the FTCF all-cause mortality outcome (two-year RD: -1.65%, 95% CI -2.32%, -0.98%).

Risk differences for the outcomes of all stroke and major bleeding are listed in **Table 5.** The all stroke outcome looked similar to ischemic stroke outcome, though farther from the null in AA analyses (two-year IPTW RD: -0.94%, 95% CI -1.63%, -0.25%, SMR RD: -0.86%, 95% CI -1.34%, -0.38%), while the major bleeding outcome showed increased risk of bleed in the IPTW AA analyses (two-year RD: 1.04%,

95% CI -1.06%, 3.15%) but no real difference in the SMR analyses (two-year RD: -0.18%, 95% CI -1.10%, 0.74%); both estimates were fairly imprecise, however.

Changing the allowable gap in medication supply to 7 days resulted in less harmful RDs for gastrointestinal bleeding in the IPTW AA analyses (two-year RD: 0.08%, 95% CI -1.19%, 1.36%). A smaller attenuation was observed in the IPTW AA gastrointestinal bleeding RD with the 60-day gap analysis (RD 0.84%, 95% CI -0.04%, 1.73%), as the individual with a weight of 42 was no longer classified as a new user of dabigatran. Removing the capacity for procedure codes for anticoagulation management to extend treatment episodes attenuated the apparent mortality benefit in the IPTW AA analyses (RD -2.06%, 95% CI -4.45%, 0.32%). The general trend of FTCF and AA of ischemic stroke RDs on opposite sides of the null was preserved through these analyses. The exclusion of anyone with a stroke in the past six months diminished the favorable IPTW AA RD for ischemic stroke (RD -0.31%, 95% CI -1.09%, 0.5%). Restricting the population to patients with a predicted probability of frailty of less than 10% at baseline excluded 42% of patients and reduced the scale of the AA all-cause mortality RD in both the IPTW (RD -1.13%, 95% CI -2.43%, -0.33%) and SMR (RD -1.63%, 95% CI -2.54%, -0.73%) analyses; FTCF all-cause mortality RDs were largely unaffected. Full risks of each outcome and results from sensitivity analyses are listed in **Table 6 and 7.**

4.4: Discussion

There is an ongoing shift in stroke prophylaxis for atrial fibrillation patients away from warfarin and towards novel oral anticoagulants in the United States,[128] but it is difficult to capture the extent to which that is improving patient outcomes in the absence of estimates of absolute-scale treatment effects. This study is among the first to estimate absolute-scale effects of dabigatran initiation compared to warfarin initiation on all-cause mortality, ischemic stroke, and gastrointestinal bleeding in the United States population of older adults in Medicare while also comparing adherence-adjusted and first treatment carried forward estimates. In adherence-adjusted analyses, we saw decreases in all-cause mortality and risk of ischemic stroke among the dabigatran new users with a higher incidence of gastrointestinal bleeding. On the other hand, first treatment carried forward analyses showed attenuation of the estimated mortality benefits and reversal of the estimated stroke benefits. Estimated dabigatran benefits were slightly greater and estimated harms slightly less in the dabigatran new users than in the entire population of new users.

The adherence-adjusted results of this study generally mirror those of other studies that estimated relative scale treatment effects and censored at treatment discontinuation, which had found hazard ratios between 0.73 and 0.80 for ischemic stroke and between 1.04 and 1.28 for gastrointestinal bleeding where our study found slightly more extreme IPTW risk ratios of 0.70 and 1.48, respectively.[14-17,19-26] A protective association between dabigatran use and all-cause mortality has been universally observed, with hazard ratios of between 0.57 and 0.76 compared to our risk ratio of 0.78. Notably, our IPTW results for gastrointestinal bleeding (risk ratio: 1.48) align closely with the hazard ratios observed in the RE-LY trial (trial hazard ratio: 1.50), but this is partly due to one dabigatran user standing in for a large number of warfarin users in the IPTW analysis; the SMR results (risk ratio: 1.15) and many of the sensitivity analyses are attenuated and more comparable to other non-experimental studies. Additionally, the adherence-adjusted results for stroke and gastrointestinal bleeding align better

with the findings of RE-LY than the first treatment carried forward results. This is not surprising given the better persistence in the trial (85% of dabigatran users and 90% of warfarin users on treatment after 1 year) than in this study (29% of the dabigatran users and 45% of the warfarin new users on treatment after 1 year). This suboptimal treatment persistence after initiation is not unique to this study; in an FDA analysis conducted in Medicare with a shorter gap period only half of dabigatran and warfarin users filled more than 1 prescription for an oral anticoagulant.[19]

Divergence between the first treatment carried forward and adherence-adjusted estimates is the most interesting finding, particularly for ischemic stroke estimates in all treated patients where the first treatment carried forward RD is equal in magnitude to the adherence adjusted but on the opposite side of the null. This result persisted across a variety of sensitivity analyses. This, along with the fact that median time on treatment was much higher in warfarin than dabigatran patients with more patients continuing on treatment, suggests that there may be issues with treatment persistence and adherence in the dabigatran users resulting in suboptimal stroke and mortality outcomes; this could be an artifact of the way we classified time on treatment, however. Past analyses of trial data and use data from the Veteran Affairs medical system also suggest that stroke risk may be elevated immediately after discontinuation of novel oral anticoagulants.[129,130] In our study, increased ischemic stroke risk after switching or a gap in treatment in dabigatran users was elevated relative to those who stayed on treatment but without any short-term spikes (even in the 7-day gap analyses); this is likely due to a lack of data surrounding exact time of stopping or switching. This may be a target for future research estimating per-protocol treatment effects using medical record or lab data that can take into account when treatment discontinuation or switching to another medication is clinically appropriate rather than motivated by cost concerns or falling out of clinical care.[107]

Another point worth discussing is the magnitude of the protective association between treatment with dabigatran versus warfarin and all-cause mortality in both the first treatment carried

forward and adherence-adjusted analyses. Putting these benefits on the absolute scale gives some important context: if these estimates are unbiased, they suggest an adherence-adjusted two-year mortality number-needed-to-treat (NNT) with dabigatran to prevent one death of 34 in the general new user population and a first treatment carried forward two-year mortality NNT of 57 in dabigatran users. As mentioned before, this protective association has also been observed in other studies specifically in older adults.[131] On the other hand, mortality as an outcome may be subject to more unmeasured confounding factors than those measured in claims data, particularly socioeconomic status or frailty, resulting in an exaggerated treatment benefit across these and other studies.[132-134] Reduced mortality benefits in the analyses restricted to those with less than 10% predicted probability of frailty adds some support to this hypothesis, though some of this may be due to lower overall risks in those populations.

This study has several limitations to consider. Variables associated with treatment initiation and the outcomes that we were not able to capture in claims data could bias treatment effect estimates. In particular, we did not have data on socioeconomic status. Since the generally higher direct cost to patients of dabigatran therapy[135] may be associated with decreased use among those with lower socioeconomic status, if greater wealth improves health above and beyond wealth's influence on the measured confounders, the estimates would overestimate the benefit and underestimate the harm of dabigatran provided no other confounding exists. We also did not use race in the main analyses because Medicare's race variables perform poorly for some minority groups and, if our causal diagrams are correct, frailty, age, and comorbidity should be sufficient to block the paths from race to outcome.[136,137] When we did include Medicare's race variable (split into white, African-American, and other categories) in a post-hoc analysis, 91.5% of new users were white, and risk differences shifted at most by seven hundredths of a percent; our findings may not generalize to minority populations. Additionally, there could be unmeasured variables associated with treatment switching and discontinuation; this would lead to remaining selection bias in the adherence-adjusted estimates.

Outcomes may have been misclassified; for example, hemorrhagic strokes might be recorded as ischemic strokes or vice versa. The fact that stroke and bleeding outcomes required diagnoses after hospitalization also means that catastrophic strokes that cause death before assistance can arrive or in the emergency department might not be captured, which could explain the mortality benefit above and beyond stroke prevention. Finally, differential measurement error in exposure due to differing grace periods or patterns of use for the two drugs to take into account warfarin dosage changes could result in biased adherence-adjusted estimates, as one drug may have more time incorrectly classified as on-treatment than the other. That would likely make that drug look comparatively safer, but less efficacious. That said, sensitivity analyses using alternative grace periods showed generally similar results for the mortality and ischemic stroke outcomes, and our adherence-adjusted results were overall similar to past non-experimental work with differing grace periods that ignored procedure codes. Finally, our results may not apply to younger populations, populations outside the United States, or Medicare beneficiaries that enroll in managed care; this is especially our first treatment carried forward estimates, where differing cost profiles may result in substantially differing persistence and adherence.

These limitations aside, the absolute effect estimates from this study represent a step forward in understanding oral anticoagulant treatment in older adults with atrial fibrillation. More and more patients (and more and more older adults) with atrial fibrillation are taking novel oral anticoagulants. Based on these findings, dabigatran may be a better choice than warfarin in this population as a whole when they remain on treatment; lower treatment persistence in the dabigatran users may mean these benefits are not being realized in the real world, however. As researchers accumulate data on these people, estimating risk and rate differences in addition to hazard ratios and relative risks is critical if we want to identify the best treatments for older adults, particularly when we believe treatments may help some patients but harm others. Additionally, while it is tempting to conduct only one of an first treatment carried forward or adherence adjusted analysis in a study, implementing both can identify

how patterns in adherence can shape the potential benefits and harms of treatments, especially in settings where the two treatments may differ in patterns of persistence.[138,139] Here, conducting both first treatment carried forward and adherence-adjusted analyses allowed us to identify differences in time on treatment for the two medications that could result in less than optimal outcomes for initiators of dabigatran. Further exploration of the effects of both traditional and novel oral anticoagulant adherence and persistence that take into account potential bias from treatment switching and discontinuation are key to understanding the comparative safety and effectiveness of oral anticoagulants in older adults.

Figure 16: Study flow diagram showing study inclusion and exclusion of new use periods identified in the Medicare 20% random sample.

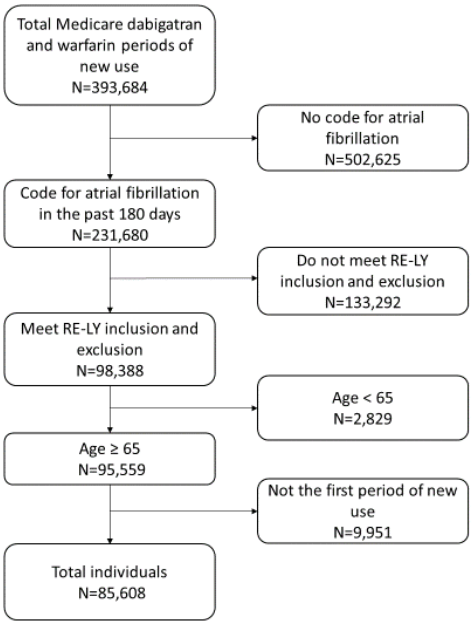

Table 2: Distributions of Covariates Included in the Propensity Score Model in New Users of Warfarin and Dabigatran in a Medicare Population

Covariate	Dabigatran new users N=10,717 (%)		Warfarin new users N=74,891 (%)		Absolute Standardized Mean Difference (ASMD)
Age					
65-69	2,249	21.0%	10,295	13.7%	0.192
70-74	2,859	26.7%	15,442	20.6%	0.143
75-79	2,514	23.5%	16,666	22.3%	0.029
80+	3,095	28.9%	32,488	43.4%	0.305
Female	5316	49.6%	32430	43.3%	0.127
Hypertension	10522	98.2%	73340	97.9%	0.018
Diabetes	3334	31.1%	24329	32.5%	0.030
Coronary Artery Disease	5178	48.3%	37389	49.9%	0.032
Congestive Heart Failure	3839	35.8%	30404	40.6%	0.098
Peripheral Vascular Disease	1626	15.2%	14829	19.8%	0.122
Past Stroke	2522	23.5%	19768	26.4%	0.066
Past TIA	900	8.4%	6276	8.4%	0.001
Hyperlipidemia	8973	83.7%	59521	79.5%	0.110
Atherosclerosis	4883	45.6%	35535	47.4%	0.038
Obesity	1348	12.6%	7960	10.6%	0.061
Smoking	739	6.9%	5149	6.9%	0.001
Cancer	1833	17.1%	11836	15.8%	0.035
Past Bleed	905	8.4%	9595	12.8%	0.142
Past Gastrointestinal Bleeding	537	5.0%	4863	6.5%	0.064
Acute Renal Dysfunction in the Past Year	317	3.0%	4303	5.7%	0.137
Alcohol Abuse	99	0.9%	677	0.9%	0.002
Ablation in the Last Year	209	2.0%	693	0.9%	0.086
Cardioversion in the Last Year	990	9.2%	3092	4.1%	0.206
Deep Vein Thrombosis	355	3.3%	8465	11.3%	0.311
Pulmonary Embolism	98	0.9%	4115	5.5%	0.262
Previous Warfarin Use (Ever)	2154	20.1%	26725	35.7%	0.353
Frailty Probability (median, P25-P75)	0.050	(0.027 – 0.114)	0.073	(0.035 – 0.225)	0.275
AA Follow-up (median, P25-P75)	152	(60-382)	259	(117 – 625)	NA
FTFC Follow-up (median, P25-P75)	980	(489-1,386)	846	(355-1,415)	NA

AA=adherence-adjusted. FTFC=first treatment carried forward. P25=25th percentile. P75 = 75th percentile.

Figure 17: Standardized mean differences in new users in the crude (empty circles), after weighting with IPTW (filled squares), and after SMR weights (filled triangles) for covariates with standardized mean differences larger than 0.10 in the crude. All measured confounders had absolute standardized mean difference smaller than 0.10 after weighting, signifying balance in these covariates.

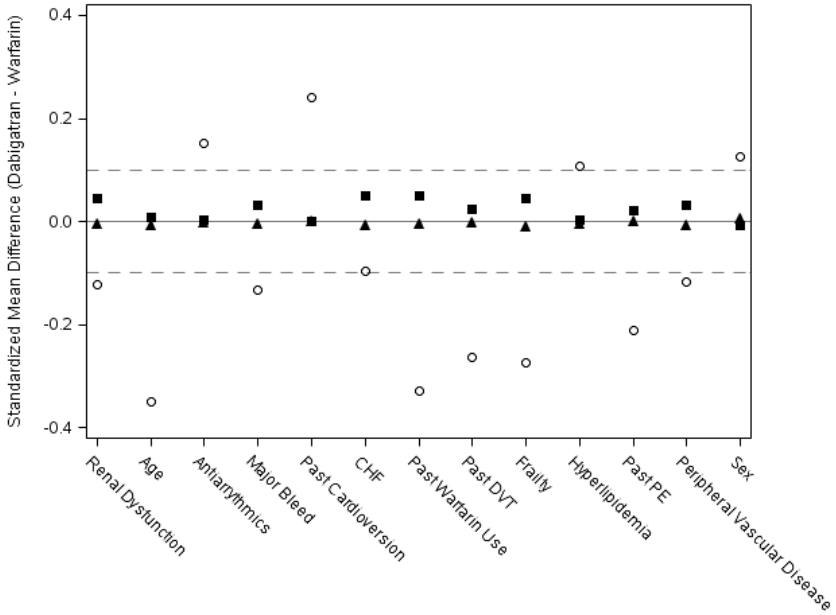

Table 3: Number of Events, Person-Years of Follow-Up, and Risks at One and Two Years By Treatment Group, Weighting Methodology, and Type of Follow-Up

Group	Person-years	Events	Incidence rate per 100 person-years	One-year Risk	Two-year Risk
Ischemic Stroke					
Dabigatran new users					
Crude, FTFC	17,259	237	1.37	1.38%	2.57%
Crude, AA	7,474	68	0.91	0.86%	1.29%
IPTW[a], FTFC	17,113	314	1.84	1.94%	3.29%
IPTW[a], AA[b]	7,575	108	1.43	1.31%	1.68%
Warfarin new users					
Crude, FTFC	112,263	1,863	1.66	1.75%	2.91%
Crude, AA	70,414	1,095	1.56	1.54%	2.48%
IPTW[a], FTFC	112,676	1,823	1.62	1.72%	2.85%
IPTW[a], AA[b]	70,670	1,068	1.51	1.51%	2.41%
SMR[a] weighted to Dabigatran, FTFC	16,539	221	1.34	1.47%	2.41%
SMR[a] weighted to Dabigatran, AA[b]	10,260	127	1.24	1.27%	1.96%
All-Cause Mortality					
Dabigatran new users					
Crude, FTFC	17,438	934	5.36	5.05%	10.24%
Crude, AA	7,493	269	3.59	3.92%	6.82%
IPTW[a], FTFC	17,343	1,476	8.51	8.29%	15.65%
IPTW[a], AA[b]	7,599	484	6.37	6.37%	10.51%
Warfarin new users					
Crude, FTFC	113,607	10,807	9.51	9.37%	17.14%
Crude, AA	70,918	5,732	8.08	8.18%	14.02%
IPTW[a], FTFC	114,005	10,385	9.11	8.97%	16.49%
IPTW[a], AA[b]	71,167	5,653	7.94	7.85%	13.49%
SMR[a] weighted to Dabigatran, FTFC	16,714	1,064	6.36	6.24%	11.89%
SMR[a] weighted to Dabigatran, AA[b]	10,326	570	5.52	5.47%	9.78%
Gastrointestinal Bleeding					
Dabigatran new users					
Crude, FTFC	17,133	330	1.93	1.91%	3.55%
Crude, AA	7,461	168	2.25	2.18%	4.00%
IPTW[a], FTFC	16,952	438	2.59	2.63%	4.55%
IPTW[a], AA[b]	7,560	272	3.60	3.42%	5.51%
Warfarin new users					
Crude, FTFC	111,617	2,288	2.05	2.21%	3.54%
Crude, AA	70,329	1,597	2.27	2.27%	3.76%
IPTW[a], FTFC	112,028	2,259	2.02	2.18%	3.50%
IPTW[a], AA[b]	70,574	1,621	2.30	2.24%	3.72%
SMR[a] weighted to Dabigatran, FTFC	16,445	294	1.79	1.97%	3.19%
SMR[a] weighted to Dabigatran, AA[b]	10,247	215	2.10	2.07%	3.49%

AA=adherence-adjusted. FTFC=first treatment carried forward. IPTW = inverse probability of treatment weighted. SMR=standardized morbidity ratio weighted.

[a]Weighted based upon logistic regression including frailty and age modeled with restricted cubic splines at the 20th, 40th, 60th, and 80th percentile, sex, diabetes, coronary artery disease, congestive heart failure, hypertension, stroke in the past year, transient ischemic attack, cancer, bleed and GI bleed history, past use of vitamin K antagonists, alcohol abuse, acute renal problems, atherosclerosis, cardioversion in the past year, deep vein thrombosis in the past year, hyperlipidemia, obesity, pulmonary embolism in the past year, peripheral vascular disease and codes indicating smoking.

[b]Weighted adherence-adjusted analyses include time-varying inverse probability of censoring weights to account for differential censoring and switching across treatment arms by measured variables.

Figure 18: Inverse probability of treatment weighted survival curves for ischemic stroke comparing the adherence-adjusted (panel A) and first treatment carried forward (panel B) follow-up schemes.

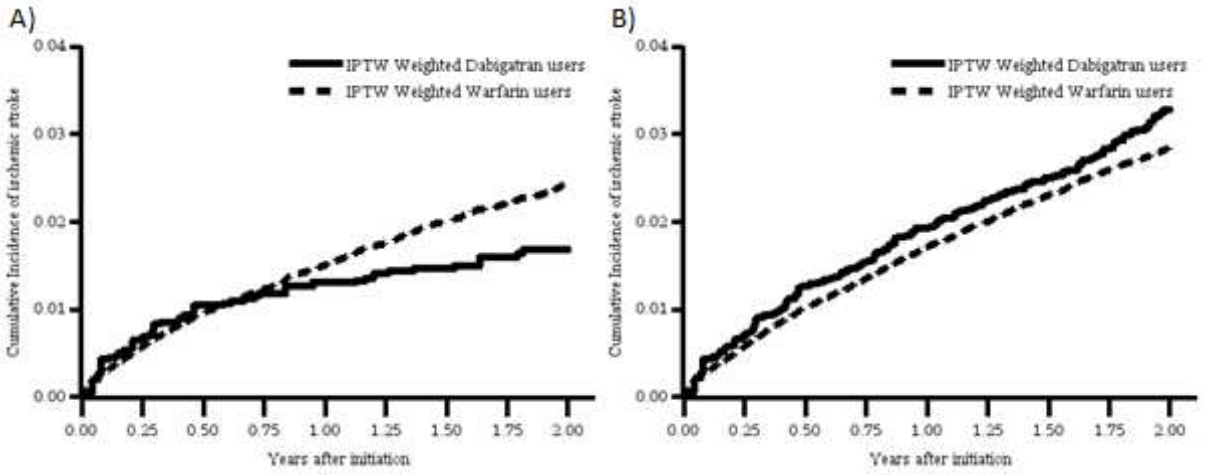

Figure 19: Inverse probability of treatment weighted survival curves for all-cause mortality comparing the adherence-adjusted (panel A) and first treatment carried forward (panel B) follow-up schemes.

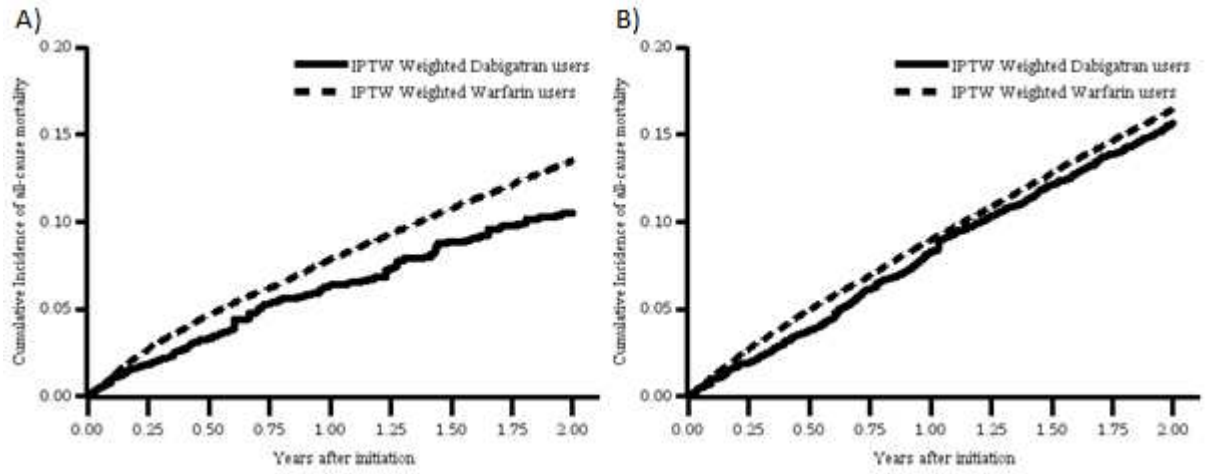

Figure 20: Inverse probability of treatment weighted survival curves for gastrointestinal bleeding comparing the adherence-adjusted (panel A) and first treatment carried forward (panel B) follow-up schemes.

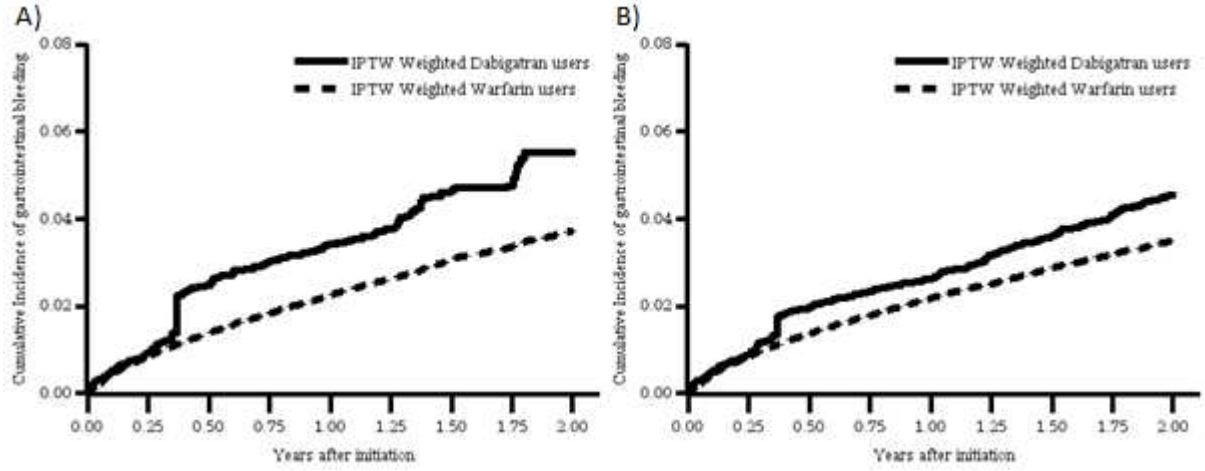

Figure 21: Inverse probability of treatment weighted survival curves for all stroke comparing the adherence-adjusted (panel A) and first treatment carried forward (panel B) follow-up schemes.

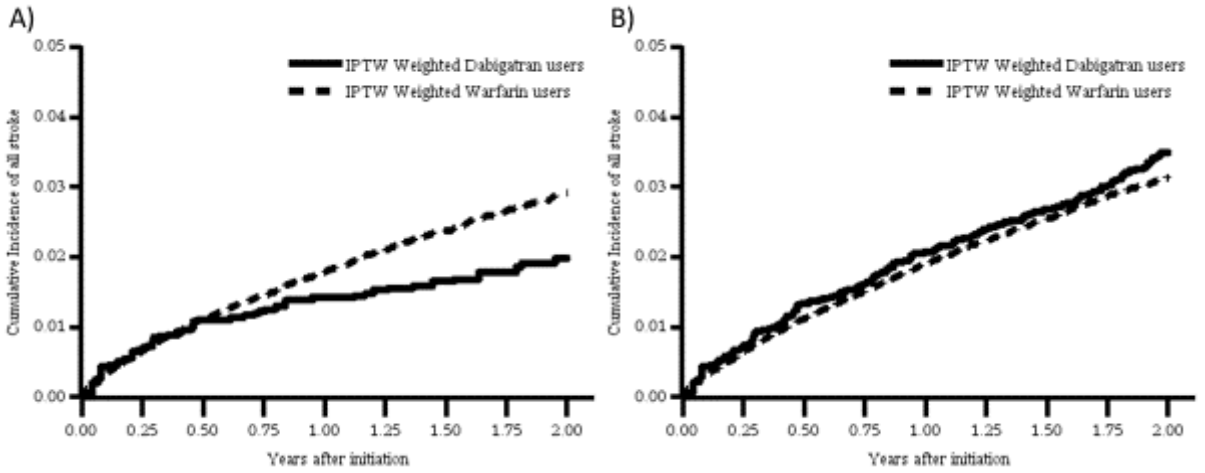

Figure 22: Inverse probability of treatment weighted survival curves for major bleeding comparing the adherence-adjusted (panel A) and first treatment carried forward (panel B) follow-up schemes.

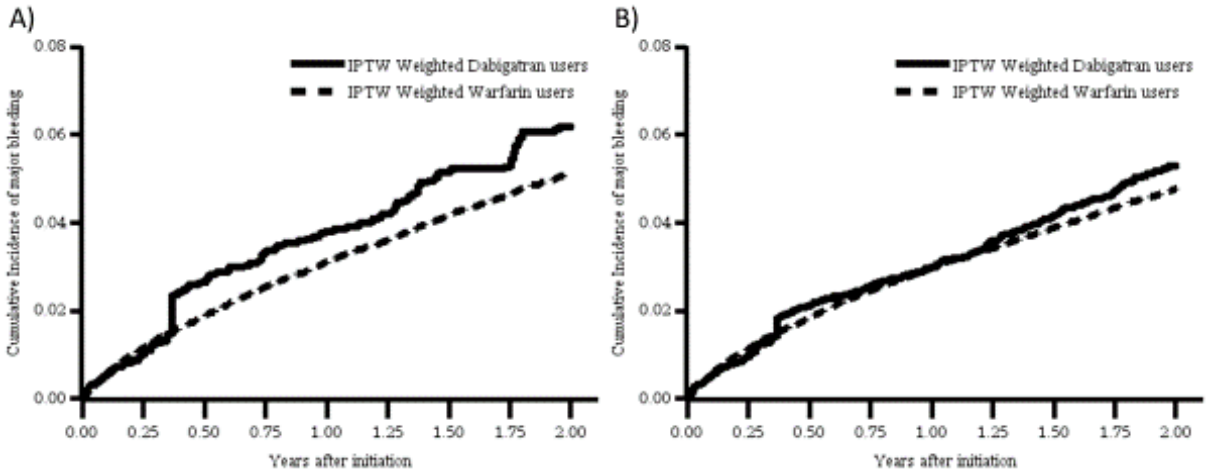

Table 4: Risk Ratios and Risk Differences Comparing Dabigatran New Users to Warfarin New Users for Two-Year Risks By Outcome, Weighting Method, and Type of Follow-Up

	Two-year adherence-adjusted[b]:		Two-year first treatment carried forward:	
Estimate	Risk ratio (95% CI)	Risk difference (95% CI)	Risk ratio (95% CI)	Risk difference (95% CI)
Ischemic Stroke				
Crude	0.52 (0.37, 0.74)	-1.19% (-1.67%, -0.71%)	0.88 (0.77, 1.02)	-0.34% (-0.71%, 0.02%)
IPTW[a]	0.70 (0.49, 1.03)	-0.73% (-1.40%, -0.06%)	1.15 (0.94, 1.41)	0.44% (-0.22%, 1.09%)
SMR[a] weighted to dabigatran	0.66 (0.48, 0.91)	-0.67% (-1.10%, -0.24%)	1.07 (0.93, 1.23)	0.16% (-0.20%, 0.52%)
All-cause Mortality				
Crude	0.49 (0.42, 0.56)	-7.20% (-8.13%, -6.27%)	0.60 (0.56, 0.63)	-6.90% (-7.59%, -6.21%)
IPTW[a]	0.78 (0.64, 0.95)	-2.98% (-5.05%, -0.91%)	0.95 (0.86, 1.04)	-0.84% (-2.39%, 0.72%)
SMR weighted to dabigatran	0.70 (0.60, 0.81)	-2.96% (-3.97%, -1.95%)	0.86 (0.81, 0.91)	-1.65% (-2.32%, -0.98%)
Gastrointestinal Bleeding				
Crude	1.06 (0.87, 1.30)	0.24% (-0.49%, 0.98%)	1.00 (0.88, 1.13)	-0.00% (-0.43%, 0.44%)
IPTW[a]	1.48 (1.05, 2.09)	1.79% (-0.13%, 3.71%)	1.30 (1.05, 1.61)	1.05% (0.08%, 2.01%)
SMR[a] weighted to dabigatran	1.15 (0.93, 1.41)	0.51% (-0.30%, 1.31%)	1.11 (0.98, 1.26)	0.36% (-0.08%, 0.79%)

SMR = standardized morbidity ratio. IPTW = inverse probability of treatment weights.

[a]Weighted based upon logistic regression including frailty and age modeled with restricted cubic splines at the 20th, 40th, 60th, and 80th percentile, sex, diabetes, coronary artery disease, congestive heart failure, hypertension, stroke in the past year, transient ischemic attack, cancer, bleed and GI bleed history, past use of vitamin K antagonists, alcohol abuse, acute renal problems, atherosclerosis, cardioversion in the past year, deep vein thrombosis in the past year, hyperlipidemia, obesity, pulmonary embolism in the past year, peripheral vascular disease and codes indicating smoking.

[b]Weighted adherence-adjusted analyses include time-varying inverse probability of censoring weights to account for differential censoring and switching across treatment arms by measured variables.

Table 5: Risk Ratios and Risk Differences Comparing Dabigatran New Users to Warfarin New Users for Two-Year Risks of Stroke and Major Bleeding By Outcome, Weighting Method, and Type of Follow-Up

	Two-year adherence-adjusted[b]:		Two-year first treatment carried forward:	
Estimate	Risk ratio (95% CI)	Risk difference (95% CI)	Risk ratio (95% CI)	Risk difference (95% CI)
All Stroke				
Crude	0.53 (0.39, 0.72)	-1.4% (-1.93%, -0.87%)	0.85 (0.75, 0.98)	-0.47% (-0.85%, -0.09%)
IPTW[a]	0.68 (0.48, 0.96)	-0.94% (-1.63%, -0.25%)	1.11 (0.93, 1.34)	0.35% (-0.31%, 1.01%)
SMR[a] weighted to dabigatran	0.65 (0.48, 0.87)	-0.86% (-1.34%, -0.38%)	1.03 (0.9, 1.18)	0.09% (-0.29%, 0.46%)
Major Bleeding				
Crude	0.89 (0.73, 1.10)	-0.64% (-1.14%, -0.15%)	0.87 (0.77, 0.97)	-0.54% (-1.43%, 0.34%)
IPTW[a]	1.20 (0.86, 1.68)	1.04% (-1.06%, 3.15%)	1.11 (0.91, 1.35)	0.53% (-0.52%, 1.58%)
SMR[a] weighted to dabigatran	0.96 (0.79, 1.18)	-0.18% (-1.10%, 0.74%)	0.97 (0.86, 1.09)	-0.14% (-0.65%, 0.36%)

SMR = standardized morbidity ratio. IPTW = inverse probability of treatment weights.

[a]Weighted based upon logistic regression including frailty and age modeled with restricted cubic splines at the 20th, 40th, 60th, and 80th percentile, sex, diabetes, coronary artery disease, congestive heart failure, hypertension, stroke in the past year, transient ischemic attack, cancer, bleed and GI bleed history, past use of vitamin K antagonists, alcohol abuse, acute renal problems, atherosclerosis, cardioversion in the past year, deep vein thrombosis in the past year, hyperlipidemia, obesity, pulmonary embolism in the past year, peripheral vascular disease and codes indicating smoking.

[b]Weighted adherence-adjusted analyses include time-varying inverse probability of censoring weights to account for differential censoring and switching across treatment arms by measured variables.

Table 6: Two-year Risks Across Sensitivity Analyses By Outcome, Treatment Group, Weighting Methodology, and Type of Follow-Up

Group	Two-year risk (7 Day gap)	Two-year risk (60 day gap)	Two-year risk (ignoring management procedure codes)	Two-year risk (no stroke in the past six months)	Two-year risk (low predicted probability of frailty)
Ischemic Stroke					
Dabigatran new users					
Crude, FTFC	2.52%	2.49%	2.57%	2.31%	1.92%
Crude, AA	1.41%	1.48%	1.30%	1.16%	0.91%
IPTW[a], FTFC	3.35%	3.09%	3.44%	2.87%	2.37%
IPTW[a], AA[b]	1.70%	2.08%	1.76%	1.64%	1.04%
Warfarin new users					
Crude, FTFC	2.88%	2.92%	2.86%	2.41%	2.01%
Crude, AA	2.33%	2.54%	2.39%	2.02%	1.64%
IPTW[a], FTFC	2.83%	2.85%	2.81%	2.35%	1.96%
IPTW[a], AA[b]	2.28%	2.47%	2.34%	1.95%	1.60%
SMR[a] weighted to Dabigatran, FTFC	2.39%	2.39%	2.38%	1.95%	1.72%
SMR[a] weighted to Dabigatran, AA[b]	1.84%	1.99%	1.93%	1.53%	1.34%
All-Cause Mortality					
Dabigatran new users					
Crude, FTFC	10.32%	9.98%	10.20%	9.24%	5.56%
Crude, AA	5.79%	7.45%	6.72%	6.01%	3.85%
IPTW[a], FTFC	16.22%	14.53%	16.36%	13.18%	6.33%
IPTW[a], AA[b]	8.81%	10.32%	10.98%	8.49%	4.74%
Warfarin new users					
Crude, FTFC	17.28%	17.00%	16.84%	15.66%	7.67%
Crude, AA	12.25%	14.57%	13.51%	12.91%	5.94%
IPTW[a], FTFC	16.68%	16.36%	16.26%	15.07%	7.52%
IPTW[a], AA[b]	11.82%	14.03%	13.04%	12.43%	5.87%
SMR[a] weighted to Dabigatran, FTFC	12.01%	11.74%	11.81%	10.99%	6.66%
SMR[a] weighted to Dabigatran, AA[b]	8.52%	10.16%	9.48%	9.14%	5.48%
Gastrointestinal Bleeding					
Dabigatran new users					
Crude, FTFC	3.46%	3.57%	3.56%	3.35%	2.81%
Crude, AA	4.04%	3.81%	3.98%	3.71%	3.02%
IPTW[a], FTFC	4.34%	4.23%	4.67%	3.94%	3.13%
IPTW[a], AA[b]	3.70%	4.56%	5.64%	4.36%	3.25%
Warfarin new users					
Crude, FTFC	3.48%	3.56%	3.49%	3.43%	2.82%
Crude, AA	3.63%	3.75%	3.65%	3.67%	2.98%
IPTW[a], FTFC	3.45%	3.51%	3.45%	3.38%	2.81%
IPTW[a], AA[b]	3.62%	3.72%	3.62%	3.63%	2.99%
SMR[a] weighted to Dabigatran, FTFC	3.19%	3.18%	3.18%	3.06%	2.71%
SMR[a] weighted to Dabigatran, AA[b]	3.50%	3.44%	3.40%	3.38%	3.03%

SMR = standardized morbidity ratio. IPTW = inverse probability of treatment weights. AA = adherence-adjusted. FTFC = first treatment carried forward.

[a]Weighted based upon logistic regression including frailty and age modeled with restricted cubic splines at the 20th, 40th, 60th, and 80th percentile, sex, diabetes, coronary artery disease, congestive heart failure, hypertension, stroke in the past year, transient ischemic attack, cancer, bleed and GI bleed history, past use of vitamin K antagonists, alcohol abuse, acute renal problems, atherosclerosis, cardioversion in the past year, deep vein thrombosis in the past year, hyperlipidemia, obesity, pulmonary embolism in the past year, peripheral vascular disease and codes indicating smoking.

[b]Weighted adherence-adjusted analyses include time-varying inverse probability of censoring weights to account for differential censoring and switching across treatment arms by measured variables.

Table 7: Two-year Risk Differences Across Sensitivity Analyses By Outcome, Weighting Methodology, and Type of Follow-Up

Group	2-year RD (7 Day gap)	2-year RD (60 day gap)	2-year RD (ignoring procedure codes)	2-year RD (no stroke in the past six months)	2-year RD (low predicted probability of frailty)
Ischemic Stroke					
Crude, FTFC	-0.36% (-0.69%, -0.03%)	-0.43% (-0.81%, -0.04%)	-0.29% (-0.61%, 0.03%)	-0.10% (-0.45%, 0.25%)	-0.09% (-0.44%, 0.26%)
Crude, AA	-0.93% (-1.64%, -0.22%)	-1.06% (-1.46%, -0.67%)	-1.09% (-1.55%, -0.64%)	-0.86% (-1.31%, -0.41%)	-0.74% (-1.20%, -0.27%)
IPTW[a], FTFC	0.53% (-0.11%, 1.17%)	0.24% (-0.29%, 0.76%)	0.63% (0.02%, 1.24%)	0.52% (-0.06%, 1.11%)	0.41% (-0.18%, 1.00%)
IPTW[a], AA[b]	-0.58% (-1.47%, 0.32%)	-0.39% (-1.07%, 0.29%)	-0.58% (-1.24%, 0.09%)	-0.31% (-1.09%, 0.46%)	-0.56% (-1.15%, 0.02%)
SMR[a] weighted to Dabigatran, FTFC	0.13% (-0.20%, 0.46%)	0.09% (-0.29%, 0.47%)	0.19% (-0.15%, 0.53%)	0.36% (0.02%, 0.70%)	0.20% (-0.16%, 0.57%)
SMR[a] weighted to Dabigatran, AA[b]	-0.44% (-1.07%, 0.2%)	-0.52% (-0.91%, -0.13%)	-0.63% (-1.07%, -0.18%)	-0.37% (-0.81%, 0.07%)	-0.44% (-0.83%, -0.04%)
All-Cause Mortality					
Crude, FTFC	-6.95% (-7.59%, -6.32%)	-7.02% (-7.72%, -6.33%)	-6.64% (-7.38%, -5.9%)	-6.42% (-7.15%, -5.70%)	-2.11% (-2.77%, -1.45%)
Crude, AA	-6.46% (-8.05%, -4.87%)	-7.12% (-8.03%, -6.22%)	-6.78% (-7.74%, -5.83%)	-6.9% (-7.86%, -5.93%)	-2.09% (-2.92%, -1.26%)
IPTW[a], FTFC	-0.46% (-1.96%, 1.04%)	-1.83% (-3.09%, -0.56%)	0.10% (-1.66%, 1.85%)	-1.89% (-3.14%, -0.63%)	-1.18% (-2.03%, -0.33%)
IPTW[a], AA[b]	-3.01% (-6.46%, 0.44%)	-3.71% (-5.37%, -2.04%)	-2.06% (-4.45%, 0.32%)	-3.95% (-5.87%, -2.02%)	-1.13% (-2.43%, 0.16%)
SMR[a] weighted to Dabigatran, FTFC	-1.69% (-2.31%, -1.07%)	-1.76% (-2.45%, -1.07%)	-1.61% (-2.34%, -0.88%)	-1.75% (-2.48%, -1.02%)	-1.11% (-1.79%, -0.42%)
SMR[a] weighted to Dabigatran, AA[b]	-2.74% (-4.42%, -1.06%)	-2.72% (-3.71%, -1.73%)	-2.76% (-3.76%, -1.75%)	-3.12% (-4.22%, -2.03%)	-1.63% (-2.54%, -0.73%)
Gastrointestinal Bleeding					
Crude, FTFC	-0.02% (-0.41%, 0.36%)	0.01% (-0.45%, 0.47%)	0.07% (-0.34%, 0.47%)	-0.08% (-0.48%, 0.33%)	-0.01% (-0.42%, 0.40%)
Crude, AA	0.41% (-0.76%, 1.57%)	0.05% (-0.52%, 0.62%)	0.33% (-0.36%, 1.02%)	0.04% (-0.65%, 0.74%)	0.04% (-0.68%, 0.76%)
IPTW[a], FTFC	0.89% (-0.27%, 2.05%)	0.71% (0.09%, 1.34%)	1.22% (0.10%, 2.34%)	0.56% (-0.03%, 1.15%)	0.32% (-0.22%, 0.86%)
IPTW[a], AA[b]	0.08% (-1.19%, 1.36%)	0.84% (-0.04%, 1.73%)	2.01% (-0.40%, 4.43%)	0.73% (-0.42%, 1.88%)	0.26% (-0.91%, 1.44%)
SMR[a] weighted to Dabigatran, FTFC	0.26% (-0.13%, 0.66%)	0.40% (-0.06%, 0.86%)	0.38% (-0.03%, 0.79%)	0.29% (-0.13%, 0.70%)	0.10% (-0.31%, 0.52%)
SMR[a] weighted to Dabigatran, AA[b]	0.55% (-0.76%, 1.86%)	0.37% (-0.31%, 1.04%)	0.58% (-0.20%, 1.36%)	0.33% (-0.51%, 1.16%)	0.00% (-0.95%, 0.94%)

SMR = standardized morbidity ratio. IPTW = inverse probability of treatment weights. AA = adherence-adjsuted. FTFC = first treatment carried forward.

[a]Weighted based upon logistic regression including frailty and age modeled with restricted cubic splines at the 20th, 40th, 60th, and 80th percentile, sex, diabetes, coronary artery disease, congestive heart failure, hypertension, stroke in the past year, transient ischemic attack, cancer, bleed and GI bleed history, past use of vitamin K antagonists, alcohol abuse, acute renal problems, atherosclerosis, cardioversion in the past year, deep vein thrombosis in the past year, hyperlipidemia, obesity, pulmonary embolism in the past year, peripheral vascular disease and codes indicating smoking.

[b]Weighted adherence-adjusted analyses include time-varying inverse probability of censoring weights to account for differential censoring and switching across treatment arms by measured variables.

CHAPTER 5: REWEIGHTING ORANGES TO APPLES: COMPARING TRANSPORTED RE-LY TRIAL AND NON-EXPERIMENTAL ESTIMATES IN ATRIAL FIBRILLATION

5.1: Introduction

Interest in generalization and transport of estimated intervention effects from study populations to external target populations is on the rise across a variety of disciplines, including epidemiology.[87,89,140-144] Weight-based standardization has been proposed as a method for achieving external validity of internally valid treatment effect estimates.[88,145,146] Benchmarking the performance of these weighting methods against outcomes observed in a subset of the target population has also been suggested.[117,147] While papers developing methods often include an example or two alongside the theoretical basis for this work, these are often single case studies with simple target populations, and there is little in-depth discussion about the validity of the transported or generalized study results or how they fit with non-experimental evidence.

Meanwhile, since the 1990s, warfarin has been used in patients with atrial fibrillation (AF) to lower the risk of stroke.[36] Use of warfarin requires careful management and dose adjustment to keep patients within the therapeutic range given its complicated pharmacokinetics and interactions with food and other medications.[120] Since 2010, new drugs commonly called novel oral anticoagulants have entered the market in the United States and internationally, starting with dabigatran (after the Randomized Evaluation of Long-Term Anticoagulants, or RE-LY, non-inferiority trial) and followed by several others.[6,44,47]

Now that these drugs are on the market, non-experimental studies have been conducted to assess whether treatment effects in patient populations in routine care match trial estimates.[12,13,15-20] Unlike the intention-to-treat RE-LY trial analyses that estimate the effect of randomization to treatment

under the trial's observed persistence and adherence, these studies typically censor patients at treatment discontinuation; the resulting estimates (often referred to "as-treated" estimates from the trial work) are conditional on remaining on treatment.[148] Compared to warfarin, investigators conducting non-experimental studies of dabigatran have typically found larger survival benefits (HR trial: 0.88, non-experimental HRs: 0.57-0.76), decreased harms from gastrointestinal bleeding (HR trial: 1.50, non-experimental HRs: 1.04-1.28 with one at 1.85) and major bleeding (HR trial: 0.93, non-experimental HRs: 0.63-0.82, with one at 1.50), but similar hazards for ischemic stroke. That said, these direct comparisons of hazard ratios are comparing apples to oranges. The differences could result from a variety of sources, including unmeasured confounding in the non-experimental studies,[58] on-treatment versus intention-to-treat effect estimates that allow for crossover and discontinuation, or the comparatively older and sicker adults in the non-experimental studies experiencing different treatment effects.[61,124,149] Moreover, these hazard and risk ratios are not readily transformable into risk differences, the main parameter of public health interest for interventions and a prerequisite for benefit-harm assessment.[58,121,122,150]

In an attempt to disentangle these potential sources of disagreement and obtain absolute effect estimates, we conducted two studies, one using weighting methods to transport trial results to a population of trial-eligible older adults and another relying entirely on claims data from those older adults. By reweighting the trial to resemble the non-experimental target population, we can get closer to an apples-to-apples comparison. Medicare often provides data for research purposes and individual-level data from RE-LY has also been made available for research through a secure online portal. As a result, we can apply weighting methods to transport treatment effect estimates to various real-world target populations of older adults and benchmark these effect estimates and their variance against those obtained in the target. Moreover, the outcome data in Medicare allows benchmarking of the

weighted trial population against observed outcomes in the target population to assess whether key assumptions for these methods have been met.

This study, then, aims to compare on-treatment risks for ischemic stroke, all-cause mortality, and gastrointestinal bleeding in separate target populations of Medicare beneficiaries starting dabigatran versus warfarin with inverse odds of sampling weights and the RE-LY trial. These estimates will then be contrasted with estimates obtained with propensity score weighting in the Medicare population alone.

5.2: Methods

5.2.1: Parameters of interest

We were primarily interested in the causal two-year risk difference for three outcomes (ischemic stroke, all-cause mortality, and gastrointestinal bleeding) comparing initiation and persistence on dabigatran versus warfarin, allowing for the competing risk of death, in Medicare beneficiaries in routine care that met eligibility criteria for the RE-LY trial. We estimated this parameter two ways: first, we transported treatment effects from RE-LY trial participants over 65 to the target populations, using treatment and covariate data from insurance claims in Medicare (calling these estimates RD_{TP}, where P corresponds to the target population); and second, we estimated treatment effects using propensity score weighting using only the Medicare claims data (RD_{OP}).

We were also interested in supplementary parameters, including the causal two-year risk differences in the RE-LY participants over 65 without sampling weights, causal two-year risk differences for outcomes of all stroke and major bleeding in our target populations, and the two-year risks associated with these and primary parameters of interest.

5.2.2: Study Populations

Trial sample-The RE-LY trial recruited 18,113 participants to assess the non-inferiority of dabigatran versus warfarin for the outcomes of stroke and major bleeding in patients with non-valvular atrial fibrillation.[6] In addition to a diagnosis of atrial fibrillation, patients needed to meet a variety of inclusion criteria including a history of stroke, transient ischemic attack, or systemic embolism; ejection fraction less than 40%; symptomatic heart failure; age over 75; or age over 65 and at least one of diabetes, documented coronary artery disease, and hypertension requiring medical treatment. Participants were randomized to one of three therapies: 150 mg of dabigatran twice daily, 110 mg of dabigatran twice daily, or warfarin therapy administered in a systematic protocol. Baseline data on a

variety of covariates was collected for potential subgroup analysis and both temporary and permanent discontinuation of the study drug was tracked.

Because we were primarily interested in older adults, we limited analyses to those over 65 (omitting those younger trial participants with histories of stroke, TIA, or heart failure), and because the 110 mg twice daily dosage was not approved for use in the United States, we focused on contrasts between the 150 mg twice daily arm and the warfarin treatment arm. The trial data was provided by Boehringer Ingelheim through Clinical Study Data Request (url: https://www.clinicalstudydatarequest.com), a service that provides access to individual-level data from randomized controlled trials to researchers through a secure data platform managed by SAS called Clinical Trial Data Transparency.

Target Populations-We were interested in estimating treatment effects in trial-eligible Medicare beneficiaries that were new users of dabigatran or warfarin the United States. To estimate the distribution of potential covariates modifying treatment effect in this population, we used a 20% simple random sample of United States Medicare beneficiaries managed by the Centers for Medicare and Medicaid services to construct a cohort of new users of either drug, including only the 150 mg twice daily dose of dabigatran. We focused on new users from 2010 (the year dabigatran was approved) to October 2015 (the date of diagnosis codes switched from ICD-9 to ICD-10-CM). Individuals were eligible for inclusion after 365 days of continuous enrollment on parts A, B, and D. Participants had to be free of warfarin or any novel oral anticoagulant for at least 60 days before their qualifying prescription. We also required at least one diagnosis for atrial fibrillation in the 180 days before or 7 days after the qualifying prescription with follow-up starting at the date of atrial fibrillation diagnosis for those with diagnoses after qualifying prescription.

After identifying these eligible new use periods, we screened each for RE-LY trial eligibility using diagnosis codes in claims data. To enter our target population, individuals had to be either over age 75

or over 65 with evidence of at least one of hypertension, diabetes, congestive heart failure, coronary artery disease, and history of stroke or transient ischemic attack in the year before their qualifying prescription. We applied exclusions for active cancer in the past 180 days, and any of endocarditis, severe or chronic kidney failure, valve replacement, and active liver disease in the past 365 days. The full list of the codes used to assess these conditions, previously used by Seeger et al., is in **Appendix B**.[18] In our main analyses we also used a predicted probability of frailty greater than 15% from a claims-based frailty prediction algorithm as a proxy for excluding individuals with a low life expectancy, though we also explored omitting this exclusion.[118] We chose 15% to be slightly more conservative and ensure there were fewer potentially-frail individuals in our targets than the 20% cutoff used in the initial validation of the frailty score.

In total, we constructed four potential target populations from Medicare that appeared eligible for the RE-LY trial: first, new users of dabigatran with less than 15% predicted probability of frailty at treatment initiation (P1); second, new users of warfarin with less than 15% predicted probability of frailty initiation (P2); third, all new users of dabigatran (P3); and fourth, all new users of warfarin (P4). In sensitivity analyses, we also considered dabigatran or warfarin users without a code for stroke in the primary inpatient position in the past 180 days (to mimic RE-LY's exclusion of those without a severe stroke in the past 180 days), as well as the population of new users with ≤10% predicted probability of frailty.

5.2.3: Exposure

Trial sample: In the main on-treatment analyses, we used assigned treatment as exposure and followed participants for 2 years, death, the end of their study participation, or six days after recorded permanent cessation of assigned treatment, whichever came first (this was the safety interval defined in

the trial protocol). This meant that individuals who were assigned to a treatment arm but did not initiate it were included for the purposes of constructing sampling weights.

Medicare sample: we followed individuals until receipt of another type of oral anticoagulant with an allowed gap of 30 days' supply for dabigatran users and 45 days' supply for warfarin users. Different grace periods were due to concerns about frequent changes in dosage amongst warfarin users. We also allowed procedure codes for anticoagulation management **(listed in Appendix D)** to "refresh" days' supply of warfarin for 30 days under the assumption patients may pay out-of-pocket for the medication or, again, have their dose frequently adjusted.

5.2.4: Outcomes

We assessed two-year cumulative incidence of the first occurrence of five outcomes after treatment initiation: ischemic stroke, all-cause mortality, gastrointestinal bleed, all stroke, and major bleeding. In this paper we present results for the first three outcomes due to their more specific nature and greater similarity between the trial and target populations; results for the final two outcomes are included in supplemental material. We used the original adjudicated outcomes from the RE-LY trial when analyzing the trial data. When obtaining estimates of cumulative risk in the Medicare population, we relied on ICD-9 codes from the primary position in inpatient hospitalizations under the assumption that these outcomes are all severe enough to qualify for hospitalization stays. The full list of codes for each outcome, validated for positive predictive value in previous claims-based studies[109] and used in past non-experimental work, is listed in **Appendix A.**

5.2.5: Covariates

We used data collected during the conduct of the RE-LY trial on a wide set of potential modifiers of treatment effect (to estimate sampling weights) and variables associated with remaining on

treatment (to estimate censoring weights). These baseline variables included age (with those over 89 combined in one category for privacy reasons), sex, use of other medications, current smoking, diabetes, congestive heart failure, coronary artery disease, past stroke, past transient ischemic attack, and past cancer. We also used trial variables on heavy alcohol use, hypertension, major or moderate bleeding while on warfarin, and past use of warfarin in the censoring weights.

In the Medicare patients, we used diagnosis and procedure codes as well as drug prescriptions to identify the presence or absence of analogous variables, as well as even more variables for use for censoring weights and confounding control (which did include past use of warfarin during the individual's time in claims). We used 1-year lookbacks for all covariates except for past use of warfarin (which used all potential lookback time). The full list of codes and criteria for each modifier is presented in **Appendix B.** In addition to its role restricting the target populations, the predicted probability of frailty was used to construct weights for censoring and confounding control.

The set of variables we included in our sampling model was built from both the trial subgroup analyses and clinical knowledge about the variables associated with outcome. Assuming the sampling diagrams in **Figures 8 and 9** are correct, the covariates included will render sampling and outcome independent except through hypertension, heavy alcohol use, past use of warfarin, and bleeding history.[89,105] Rather than use linear age or predicted probability of frailty in our sampling and censoring models, we used semiparametric restricted quadratic splines with knots at the 20^{th}, 40^{th}, 60^{th}, and 80^{th} percentiles. The major covariates missing (heavy alcohol use, general history of major bleeding, hypertension, and past warfarin use) appeared to be captured poorly or measured differently in the trial and target populations: hypertension required medical treatment in the trial compared to our pre-specified more sensitive hypertension code set; bleeding history outside of warfarin use was not available in RE-LY; prevalence of heavy drinking in Medicare appeared vastly lower, suggesting

differences in sensitivity; and past use of warfarin could not be assessing prior to enrollment onto Medicare, while it was assessed over the life course in RE-LY.

5.2.6: Statistical Analysis

Analyses without sampling weights: Risk differences in those over 65 in the trial were estimated from cumulative incidence and survival curves in both the trial and target populations using the Aalen-Johansen estimator[119] to take into account the competing risk of death for all of our non-mortality outcomes. This estimator was unweighted when assessing the intention-to-treat effect of treatment randomization. When estimating the effect of being randomized and remaining on treatment, we accounted for differential discontinuation and switching by covariates using cumulative inverse probability of censoring weights estimated from the probability of remaining uncensored between quartiles of the censoring distribution, estimating those probabilities using multivariable logistic regression.[125] Due to data constraints, we used baseline versions of all covariates in the trial but allowed target population versions of covariates to vary over time.

Sampling weights (in the trial): Our first method for estimating the main parameters of interest was transporting the RE-LY trial results was combining the above censoring weights with inverse odds of sampling weights.[88] We were unable to directly combine individual-level RE-LY trial date with the Medicare data due to stipulations in our data sharing agreements, so these weights could not be calculated directly. Instead, we took an alternate approach involving simulating data and exporting coefficients described in detail in **Section 3.2.3.** Estimated covariate-conditional odds from the final multivariable logistic regression sampling models were used to calculate inverse odds of sampling weights stabilized to the number of overall participants in the trial. We then created a trial population with similar covariate distributions as the target population. These weights were combined with the inverse probability of censoring weights to estimate on-treatment survival curves and risk differences

for each outcome. The models were deemed successful if absolute standardized mean differences between distributions of all the covariates in the sampling model for the inverse odds of sampling weighted trial and target populations were less than 0.10.[114]

Non-experimental effect estimates: The second method for estimating our main parameters of interest was the use of standardized morbidity ratio (SMR) weights combined with inverse odds of censoring weights.[151] Based on the directed acyclic graphs relating exposure and the outcomes in the target population (**Figure 11**), crude estimates of treatment effect using the non-experimental data should be biased. To account for this confounding, we estimated the probability of initiating dabigatran rather than warfarin conditional on a set of covariates that should render treatment and outcome independent with multivariable logistic regression (these included a variety of comorbid conditions used in past non-experimental studies). These probabilities were transformed to odds, with the inverse of those odds being the SMR weights. We conducted separate analyses weighted to the dabigatran new users and the warfarin new users as potential target populations. These propensity score models were deemed successful if absolute standardized mean differences between the two groups of new users in the Medicare population were less than 0.10 after weighting. These final SMR weights were then combined with inverse probability of censoring weights for use in an Aalen-Johansen estimator.

Comparing estimates: We compared the estimated two-year risk difference for the three primary outcomes comparing dabigatran and warfarin initiators transporting trial results to the non-experimental treatment effect estimates in each target population.

Diagnostic assessment of transportability: In addition to comparing treatment effect estimates, we also assessed the extent to which the weighted survival curves differed from inverse probability of censoring-weighted survival in the target population.[117,147] We compared survival curves and calculated risk differences at 730 days. If there was a large gap between the two, then the chance our main parameters of interest are not being estimated in the transported or non-experimental results

increased: either the sampling model does not render them independent, and the effect estimate will be biased on at least one scale; the way the exposure is administered differs between the two populations, and the effect estimate may be biased; or the outcome itself is measured differently in the populations.[144] We considered the plausibility and implications of each of these potential sources of error for each outcome.

5.2.7: Sensitivity analyses

We conducted two main types of post-hoc sensitivity analyses: targeting additional populations (individuals with no stroke in the primary position of an inpatient encounter in the past 180 days, individuals with less than 10% predicted probability of frailty) and including additional or changing variables from the sampling model in the main analyses (measured past warfarin use and a modified version of hypertension requiring treatment with non-beta-blockers, as well as another model with quadratic age).

All statistical analyses were performed in SAS 9.4 for Windows (Cary, NC). This study was approved by the University of North Carolina at Chapel Hill's Institutional Review Board.

5.3: Results

Study populations: **Figure 23** is a flow diagram describing study inclusion and exclusion. Of the 18,113 participants included in the original RE-LY trial, there were 15,132 over the age of 65 that were used in the transportability analyses. Of these, 10,115 were randomized to the warfarin (N= 5,069) or dabigatran 150 mg twice daily arms (N=5,046) and had no missing data on covariates in any of the sampling and censoring models. In Medicare, we identified 10,717 dabigatran new users, 74,891 warfarin new users, 8,586 dabigatran new users with less than 15% predicted probability of frailty, and 50,650 warfarin new users with less than 15% predicted probability of frailty. In the main results we focus on the frailty-restricted target populations due to potential lack of positivity. Results without the frailty restriction are included in supplemental material.

Table 8 presents the distribution of potential modifiers in the trial participants randomized to the treatments of interest and the frailty-restricted target populations. **Figure 24** graphically shows the standardized mean difference for each of these potential modifiers compared to the trial. Generally, patients in the target populations were older with more cardiovascular comorbidities, including history of stroke. **Table 9** provides similar information for the populations that were not restricted by predicted probability of frailty; differences between these two groups were even greater.

Within-trial effect estimates (unweighted): Risks of mortality and ischemic stroke were lower in the on-treatment than the intention-to-treat even after applying censoring weights (see **Table 10** for two-year risks and rates). Risk differences, however, were similar in intention-to-treat and inverse probability of censoring weighted on-treatment analyses. Dabigatran remained protective for both ischemic stroke (on-treatment two-year risk difference: -0.74%, 95% CI -1.63%, -0.14%) and all-cause mortality (on-treatment two-year risk difference: -0.64%, 95% CI -1.3%, 0.1%), but harmful for gastrointestinal bleeding (on-treatment two-year risk difference: 1.37%, 95% CI 0.66%, 2.08%).

Transported trial effect estimates: The distributions of each covariate included in the sampling model in the weighted trial populations and the target populations are presented in **Table 11.** Absolute standardized mean differences for all covariates were balanced considerably better after weighting with all falling below 0.100 when contrasting the weighted trial with its corresponding target. **Table 12** presents a similar breakdown for the unrestricted target populations; the unrestricted warfarin target population still had large differences in age, sex, past stroke, and past transient ischemic attack after weighting.

Figure 25 shows cumulative incidence curves for ischemic stroke in the trial without sampling weights (left), the trial weighted to the dabigatran users (middle, P1), and the trial weighted to the warfarin users (right, P2). The cumulative incidence is higher in both weighted trial arms compared to the original trial. **Figure 26** shows cumulative incidence curves for the unrestricted target populations, and **Figures 27-30** show cumulative incidence curves for other outcomes of interest.

Table 13 presents person-time at risk, events, incidence rates, two-year risks, and two-year risk differences for ischemic stroke, all-cause mortality, and gastrointestinal bleeding when weighting to target populations P1 and P2. **Table 14** provides results for all stroke and major bleeding events, while **Table 15** shows results when weighting to P3 and P4.

After applying sampling weights, the estimated RD for ischemic stroke in P1 was about the same as the original trial (RD_{T1}: -0.78%, 95% CI -1.69%, 0.14%), and P2's RD was slightly larger in magnitude (RD_{T2}: -0.91%, 95% CI 1.93%, 0.11%). RDs for all-cause mortality changed more after applying sampling weights: while the weighted estimate in P1 was about the same as the original trial (RD_{T1}: -0.57%, 95% CI -1.83%, 0.68%), the RD for P2 was considerably attenuated (RD_{T2}: -0.16%, 95% CI -1.71%, 1.39%). Finally, the results for gastrointestinal bleeding were straightforward: across both P1 and P2, the sampling weighted RDs were slightly greater than those the original trial (RD_{T1}: 1.75%, 95% CI 0.76%, 2.74%; RD_{T2}: 1.85% (95% CI 0.65%, 3.04%).

SMR-Weighted Non-experimental Results: When the parameters of interest were estimated using claims data alone with propensity scores and SMR weights, the two-year RD_{O1} for ischemic stroke was -0.45% (95% CI -0.85%, -0.06%) and RD_{O2} was -0.55% (95% CI -1.21%, 0.12%). RD_{O1} for two-year mortality was -1.94% (95% CI -2.85%, -1.04%) and RD_{O2} was -0.40% (95% CI -2.68%, 1.89%) and RD_{O1} and RD_{O2} for gastrointestinal bleeding were 0.14% (95% CI -0.76%, 1.03%) and 0.57% (95% CI -0.70%, 1.83%), respectively. **Table 16** includes non-experimental RDs for all outcomes and target populations. Estimates of the two-year RD for ischemic stroke were similar across trial transport and non-experimental study approaches, but estimates of RDs for mortality in P1 and gastrointestinal bleeding in both P1 and P2 differed. **Figure 31** plots the non-experimental estimates and 95% confidence limits on the horizontal axis and the transported trial estimates and 95% confidence limits on the vertical axis, with each point representing a different target population and each panel representing a different outcome (**Figure 32** provides an analog for the unrestricted target populations).

Assessing Transportability: Two-year risk differences between the weighted trial populations and the target populations of less-frail dabigatran and warfarin users for all five outcomes are presented in **Table 17** (a version for the unrestricted populations can be found in **Table 18**). Weighting appeared to put the trial dabigatran users at slightly higher risk of ischemic stroke than the Medicare dabigatran users (RD for Medicare vs. weighted trial: -0.57%, 95% CI -1.20%, 0.05%) and trial warfarin users at slightly higher risk of ischemic stroke than Medicare warfarin users (RD: -0.69%, 95% CI -1.48%, 0.09%), with similar results for all stroke.

Risk of gastrointestinal bleeding also appeared higher in the weighted trial than Medicare dabigatran users (RD: -0.90%, 95% CI 2.05%, 0.25%), though it was about the same when weighting to Medicare warfarin users (RD: 0.36%, 95% CI -0.41%, 1.13%). Major bleeding was much lower in both target populations than the corresponding weighted trial (RD for dabigatran users: -4.76%, 95% CI -6.23%, -3.29%; RD for warfarin users: -4.05%, 95% CI -5.13%, -2.98%).

Finally, Medicare dabigatran users and the weighted trial dabigatran users experienced similar mortality (RD: 0.18%, 95% CI -1.07%, 1.44%), though the trial warfarin users had a marked mortality advantage over the Medicare warfarin users (RD: 2.02%, 95% CI 0.84%, 3.20%).

Sensitivity analyses: **Table 19** and **Table 20** show the results of the sensitivity analyses. Excluding those with a stroke within the past 180 days and further frailty restriction appeared to make RD_{T2} more favorable for dabigatran, particularly with respect to mortality and major bleeding; this lends further credence to the idea that mortality benefits of dabigatran (or the harms of warfarin) may vary across population. Including the modified hypertension definition and history of warfarin use in the sampling model appeared to attenuate the ischemic stroke benefits in when targeting dabigatran users, though the increase in variance was substantial. Use of a simpler quadratic age term generally left results unchanged.

5.4: Discussion

This work represents an important step in assessing the extent to which trial and non-experimental findings really disagree about treatment effects of dabigatran versus warfarin in patients with atrial fibrillation and outlines many of the difficulties encountered when trying to compare transported trials to real-world evidence in target populations of interest. Absolute-scale on-treatment effect estimates for initiating dabigatran versus warfarin were generally similar whether estimates were obtained by transporting trial results or using claims data with propensity score methods. On the other hand, transported estimates were closer to the null compared to non-experimental estimates of treatment effect on mortality for dabigatran initiators and farther from the null than non-experimental estimates of treatment effects for gastrointestinal bleeding. Weighting was able to standardize the RE-LY trial to Medicare target populations of warfarin and dabigatran new users with low predicted probability of frailty (P1 and P2), but problems standardizing to target populations that were not restricted based on frailty may reflect a lack of positivity for these older and sicker in the original randomized control trial.

Contextualizing and understanding these results and how they relate to each outcome is key, especially as our non-experimental estimates are generally in line with past non-experimental research. Perhaps the most encouraging finding from the study is that the absolute-scale effects on ischemic stroke obtained via sampling weights were equal to or greater than those from non-experimental work, especially as thrombotic strokes are the main reason to prescribe dabigatran and warfarin.[22] That the weighted trials absolute-scale effects of dabigatran versus warfarin on mortality are attenuated (and close to the null) when targeting less-frail warfarin users could either reflect confounding in the non-experimental estimates or different patterns of warfarin management in RE-LY compared to the general population that result in meaningfully different care.[71-74,78] Meanwhile, non-experimental treatment effect estimates for gastrointestinal bleeding being closer to the null could be due to a lack of data on

bleeding risk factors when transporting trial results or differences in treatment effects, or could be due to differing definitions of the outcome in the two studies.

Results from transportability assessment were mixed. The fact that the less-frail dabigatran users and the weighted trial dabigatran users had very similar risks of mortality at two years is reassuring. The other findings all have many potential explanations. Differences in mortality between Medicare warfarin users and weighted trial warfarin users could reflect overall better warfarin management in the trial or it could reflect differences in unaccounted for modifiers.[70] Using claims-based data for stroke outcomes can be difficult.[152] While our other outcomes were validated for positive predictive value, there is less available data on their sensitivity, which is key when juxtaposing trial and target population outcomes. The weighted trial versus target differences in ischemic stroke risk could plausibly be due to a combination of lower sensitivity for these outcomes in Medicare claims data (a more inclusive algorithm still only had 82% sensitivity)[153] and the fact that strokes or bleeds that caused death before hospitalization would not be captured in Medicare. Additional lowering of all stroke in the warfarin users in routine care may reflect that hemorrhagic strokes are less likely to occur due to underanticoagulation in routine care or that hemorrhagic strokes are even more poorly captured in claims data (75% sensitivity with a more inclusive algorithm).[153] The differences in gastrointestinal bleeding risk could reflect the fact that we could not use general bleeding history for standardization alongside a lower sensitivity, and the large differences in major bleeding most likely means the outcome codes in Medicare are not equivalent to RE-LY's adjudication method.

It's important to understand the limitations of our findings. While restricting to those eligible for the RE-LY trial was necessary to allow for positivity when transporting treatment effects, those patients still represent an important segment of the treated population; our findings may not apply to them. The codes and covariates defined using claims data were assumed to map to and reflect the clinical covariates obtained during the RE-LY study with minimal misclassification. While we attempted to put as

many covariates from the trial into the sampling model as we reasonably could, any effect measure modifiers missing from it could bias transported estimates of treatment effects. Furthermore, we were not able to include any interaction terms in our sampling model; this could be problematic if some effect measure modifiers need to both be present to result in altered treatment effect.[145] We relied on our inverse probability of censoring weights to mitigate the potential for time-varying confounding; however, it is possible these models also omitted key covariates and that residual bias remains. Finally, requiring fee-for-service Medicare coverage to observe prescription exposure to dabigatran restricts the study population and the resulting covariate distributions may differ from the ideal target populations.[154]

Overall, this is one of the first applications of inverse odds of sampling weights to understand the extent to which trial and non-experimental findings of treatment effects disagree about treatment effects in target populations of interest after taking into account treatment effect heterogeneity. As data from trials and clinical practice populations becomes more available for health research, researchers should make sure to specify target populations when considering whether treatment effect estimates agree with one another. While it may require additional data and analysis, comparing apples to apples rather than apples to oranges is key to understanding how and why trial results differ from non-experimental work.

Figure 23: Consort diagram for the study (RE-LY) and target (Medicare) populations. Subjects in A were required to match the inclusion and exclusion criteria of RE-LY. T1 and T4 represent the various target populations used in the analysis, with T1 and T2 referring to the main populations analyzed. All trial participants in B were used to calculate inverse odds of sampling weights. Population W and D were used to estimate treatment effects in the randomized trial.

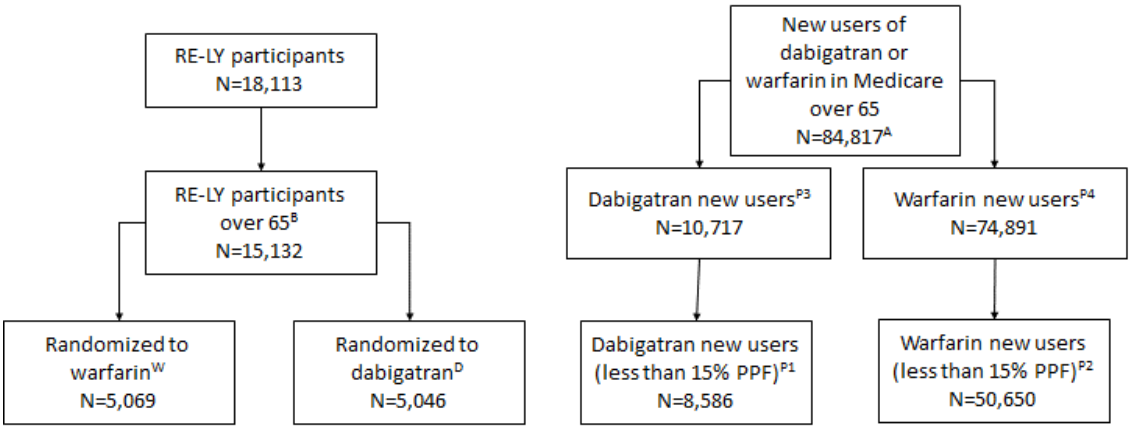

Table 8: Distribution of selected potential modifiers in RE-LY trial participants randomized to dabigatran 150 mg twice daily or warfarin and the target populations identified in the 20% random Medicare sample from 2010 through 2015 after restricting to individuals with less than 15% predicted probability of frailty.

Variable	In RE-LY and randomized to either treatment, N=10,115	Medicare dabigatran new users, N=8,586 (P1)	Medicare warfarin new users, N=50,650 (P2)
Age (Median, P25-P75)	74 (70-78)	74 (70-79)	76 (71-81)
Male Sex	6,171 (61.0%)	4,418 (51.5%)	23,691 (46.8%)
Smoking N (%)	615 (6.1%)	523 (6.1%)	3,100 (6.1%)
Past stroke N (%)	1147 (11.3%)	1,484 (17.3%)	8,630 (17.0%)
Hypertension[a] N (%)	8152 (80.6%)	8,418 (98.0%)	49,488 (97.7%)
Past TIA N (%)	871 (8.6%)	511 (6.0%)	2,575 (5.1%)
CHF N (%)	2637 (26.1%)	2,730 (31.8%)	17,519 (34.6%)
CAD N (%)	2916 (28.8%)	3,896 (45.4%)	23,171 (45.7%)
DM N (%)	2319 (22.9%)	2,506 (29.2%)	15,487 (30.6%)
Cancer N (%)	1207 (11.9%)	1,459 (17.0%)	7,784 (15.4%)
Past warfarin use[a] N (%)	6894 (68.2%)	1,565 (18.2%)	18,053 (35.6%)
Amiodarone use N (%)	994 (9.8%)	888 (10.3%)	5,020 (9.9%)
PPI use	1500 (14.8%)	2,229 (26.0%)	12,764 (25.2%)

P1=Target population 1. P2 = target population 2. CHF=Congestive heart failure. CAD = coronary artery disease. DM = diabetes mellitus. PPI = proton pump inhibitor.

[a]These variables may not be equivalent across data sources, as hypertension in the trial is treated hypertension while hypertension in the targets is treated or untreated hypertension, and past use of vitamin K antagonists in the target populations can only be examined during their time in Medicare.

Table 9: Distribution of selected potential modifiers in RE-LY trial participants randomized to dabigatran 150 mg twice daily or warfarin and the target populations identified in the 20% random Medicare sample from 2010 through 2015 that were not restricted by frailty probability.

Variable	In RE-LY and randomized to either treatment, N=10,115	Medicare dabigatran new users, N=10,717 (T3)	Medicare warfarin new users, N=74,891 (T4)
Age (Median, P25-P75)	74 (70-78)	75 (70-80)	78 (72-84)
Sex	6,171 (61.0%)	5,316 (49.6%)	32,430 (43.3%)
Smoking N (%)	615 (6.1%)	739 (6.9%)	5,149 (6.9%)
Past stroke N (%)	1147 (11.3%)	2,522 (23.5%)	19,768 (26.4%)
Hypertension[a] N (%)	8152 (80.6%)	10,522 (98.2%)	73,340 (97.9%)
Past TIA N (%)	871 (8.6%)	900 (8.4%)	6,276 (8.4%)
CHF N (%)	2637 (26.1%)	3,839 (35.8%)	30,404 (40.6%)
CAD N (%)	2916 (28.8%)	5,178 (48.3%)	37,389 (49.9%)
DM N (%)	2319 (22.9%)	3,334 (31.1%)	24,329 (32.5%)
Cancer N (%)	1207 (11.9%)	1,833 (17.1%)	11,836 (15.8%)
Past warfarin use[a] N (%)	6894 (68.2%)	2,154 (20.1%)	26,725 (35.7%)
Amiodarone use N (%)	994 (9.8%)	1,160 (10.8%)	7,606 (10.2%)
PPI use	1500 (14.8%)	3,018 (28.2%)	21,915 (29.3%)

P3=Target population 3. P4 = target population 4. CHF=Congestive heart failure. CAD = coronary artery disease. DM = diabetes mellitus. PPI = proton pump inhibitor.

[a]These variables may not be equivalent across data sources, as hypertension in the trial is treated hypertension while hypertension in the targets is treated or untreated hypertension, and past use of vitamin K antagonists in the target populations can only be examined during their time in Medicare.

Figure 24: Plot of standardized mean differences (SMDs) for covariates when comparing the RE-LY population participants over 65 to the Medicare target populations restricted to less than 15% predicted probability of frailty. Negative values indicate increased prevalence or mean in the target population compared to the trial, while positive values indicate traits that were more common in the trial than the target. The squares are SMDs between the trial and dabigatran users (P1) and the triangles are SMDs between the trial and warfarin users (P2).

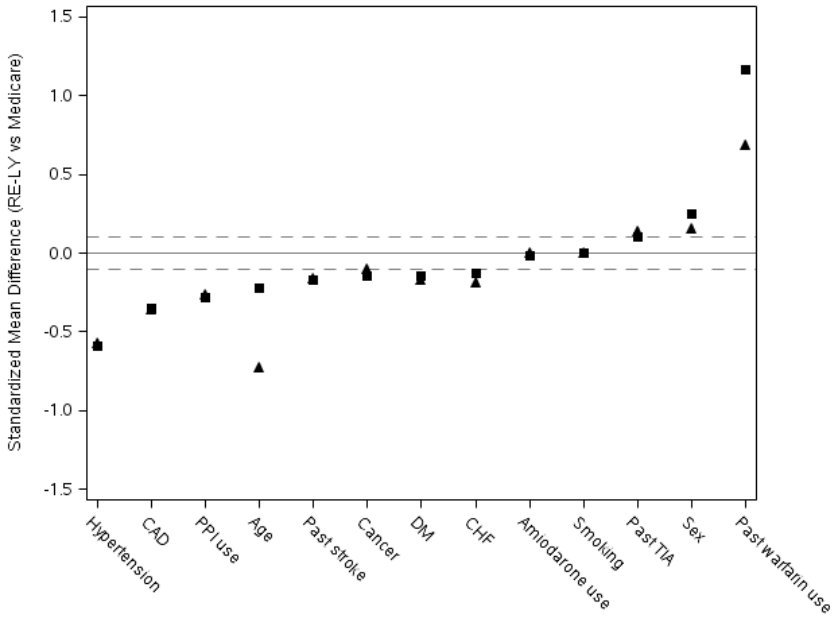

Table 10: Person-years, event numbers, two-year risks, and two-year risk differences for ischemic stroke, mortality, gastrointestinal bleeding, all stroke, and major bleeding in the intention-to-treat and as-treated analyses of the RE-LY trial participants over 65.

Outcome and treatment arm	Person-years	Events	Two-year risk	Two-year risk difference (95% CI)
Ischemic stroke-ITT				
Warfarin	8,751	720	2.39%	Ref.
Dabigatran 150 mg	8,764	650	1.75%	-0.64% (-1.3%, 0.01%)
All-cause mortality-ITT				
Warfarin	8,823	336	7.44%	Ref.
Dabigatran 150 mg	8,822	308	6.69%	-0.74% (-1.63%, 0.14%)
Gastrointestinal bleeding-ITT				
Warfarin	8,739	724	2.25%	Ref.
Dabigatran 150 mg	8,661	715	3.62%	1.37% (0.66%, 2.08%)
All stroke-ITT				
Warfarin	8,736	703	3.31%	Ref.
Dabigatran 150 mg	8,758	648	2.10%	-1.21% (-1.97%, -0.45%)
Major bleeding-ITT				
Warfarin	8,549	849	7.57%	
Dabigatran 150 mg	8,514	809	7.81%	0.24% (-0.91%, 1.38%)
Ischemic stroke-OT, IPCW				
Warfarin	7,976	1,321	2.14%	Ref.
Dabigatran 150 mg	7,513	1,103	1.33%	-0.81% (-1.48%, -0.14%)
All-cause mortality-OT, IPCW				
Warfarin	8,025	579	4.39%	
Dabigatran 150 mg	7,528	483	3.71%	-0.69% (-1.58%, 0.21%)
Gastrointestinal bleeding-OT, IPCW				
Warfarin	7,969	1,331	2.10%	Ref.
Dabigatran 150 mg	7,465	1,308	3.64%	1.54% (0.82%, 2.25%)
All stroke-OT, IPCW				
Warfarin	7,973	1,317	3.00%	Ref.
Dabigatran 150 mg	7,513	1,095	1.56%	-1.44% (-2.22%, -0.67%)
Major bleeding-OT, IPCW				
Warfarin	7,846	1,835	7.14%	Ref.
Dabigatran 150 mg	7,379	1,718	7.76%	0.62% (-0.54%, 1.77%)

CI = confidence interval. ITT=intention-to-treat. OT=on-treatment.

Table 11: Distributions of the covariates included in the sampling model in each of the four target populations in Medicare and the associated weighted trial populations in the RE-LY trial.

Variable	Dabigatran new users, N=8,586 (P1)	Trial weighted to P1, N=10,113.57	SMDs (P1 weighted trial vs P1)	Warfarin new users, N=50,650 (P2)	P2 weighted trial, N=10,073.6	SMDs (P2 weighted Trial vs P2)
Age (Median, P25-P75)	74 (70-79)	74 (70-79)	0.043	76 (71-81)	76 (71-81)	-0.008
Sex N (%)	4418 (51.5%)	5233.7 (51.7%)	0.005	23691 (46.8%)	4864.5 (48.2%)	0.029
Smoking N (%)	523 (6.1%)	589.7 (5.8%)	-0.011	3100 (6.1%)	572.5 (5.7%)	-0.018
Past stroke N (%)	1484 (17.3%)	1735.2 (17.2%)	-0.004	8630 (17%)	1697.6 (16.8%)	-0.004
Past TIA N (%)	511 (6%)	574.7 (5.7%)	-0.014	2575 (5.1%)	485.8 (4.8%)	-0.013
CHF N (%)	2730 (31.8%)	3041.7 (30.1%)	-0.037	17519 (34.6%)	3153.9 (31.3%)	-0.071
CAD N (%)	3896 (45.4%)	4581.2 (45.3%)	-0.002	23171 (45.7%)	4567.8 (45.3%)	-0.008
DM N (%)	2506 (29.2%)	2917.7 (28.8%)	-0.008	15487 (30.6%)	2979.4 (29.5%)	-0.023
Cancer N (%)	1459 (17%)	1817.9 (18%)	0.026	7784 (15.4%)	1737.6 (17.2%)	0.050
Amiodarone use N (%)	888 (10.3%)	1015 (10%)	-0.009	5020 (9.9%)	894.7 (8.9%)	-0.035
PPI use	2229 (26%)	2655.4 (26.3%)	0.006	12764 (25.2%)	2648.4 (26.3%)	0.024

P1=Target population 1. P2 = target population 2. CHF=Congestive heart failure. CAD = coronary artery disease. DM = diabetes mellitus. PPI = proton pump inhibitor.

Table 12 Distributions of the covariates included in the sampling model in each of the four target populations in Medicare and the associated weighted trial populations in the RE-LY trial.

Variable	Dabigatran new users, N=10,717 (P3)	Trial weighted to P3, N=9,991.26	SMDs (P3 weighted trial vs P3)	Warfarin new users, N=74,891 (P4)	Trial weighted to P4, N=7,980.4	SMDs (P4 Weighted Trial vs P4)
Age (Median, P25-P75)	75 (70-80)	75 (70-81)	0.061	78 (72-84)	74 (72-75)	-1.459
Sex N (%)	5316 (49.6%)	5,089.6 (50.9%)	0.027	32430 (43.3%)	4,299.9 (53.7%)	0.210
Smoking N (%)	739 (6.9%)	633.6 (6.3%)	-0.023	5149 (6.9%)	675.1 (8.4%)	0.058
Past stroke N (%)	2522 (23.5%)	2256.6 (22.6%)	-0.022	19768 (26.4%)	1754.2 (21.9%)	-0.105
Past TIA N (%)	900 (8.4%)	734 (7.3%)	-0.039	6276 (8.4%)	439.1 (5.5%)	-0.115
CHF N (%)	3839 (35.8%)	3290.1 (32.9%)	-0.061	30404 (40.6%)	2931.3 (36.6%)	-0.082
CAD N (%)	5178 (48.3%)	4746 (47.5%)	-0.016	37389 (49.9%)	3938.4 (49.2%)	-0.014
DM N (%)	3334 (31.1%)	2961.7 (29.6%)	-0.032	24329 (32.5%)	2536.6 (31.7%)	-0.017
Cancer N (%)	1833 (17.1%)	1860.8 (18.6%)	0.040	11836 (15.8%)	1332 (16.6%)	0.023
Amiodarone use N (%)	1160 (10.8%)	1030.8 (10.3%)	-0.016	7606 (10.2%)	764.6 (9.6%)	-0.022
PPI use	3018 (28.2%)	2841.1 (28.4%)	0.005	21915 (29.3%)	2207.2 (27.6%)	-0.038

P3=Target population 3. P4 = target population 4. CHF=Congestive heart failure. CAD = coronary artery disease. DM = diabetes mellitus. PPI = proton pump inhibitor.

Figure 25: Cumulative incidence of ischemic stroke during two years of follow-up and censoring after discontinuation and applying censoring weights in A) the on-treatment trial population without sampling weights, B) the trial population weighted to the less-frail initiators of dabigatran in Medicare (P1), and C) the trial population weighted to the less-frail initiators of warfarin in Medicare (P2). The solid line shows survival for dabigatran users and the dashed line corresponds to warfarin users.

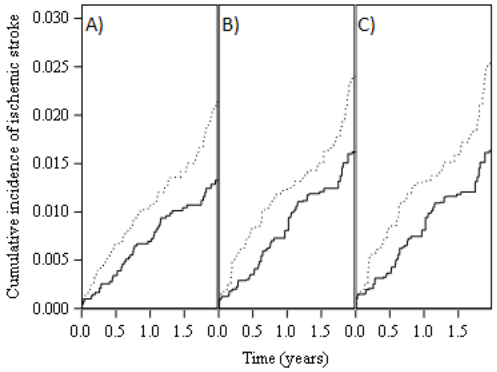

Figure 26: Cumulative incidence of ischemic stroke during two years of follow-up and censoring after discontinuation and applying censoring weights in the as-treated A) trial population without sampling weights, B) trial population weighted to initiators of dabigatran in Medicare (P3), and C) trial population weighted to initiators of warfarin in Medicare (P4). The solid line shows survival for dabigatran users and the dashed line corresponds to warfarin users.

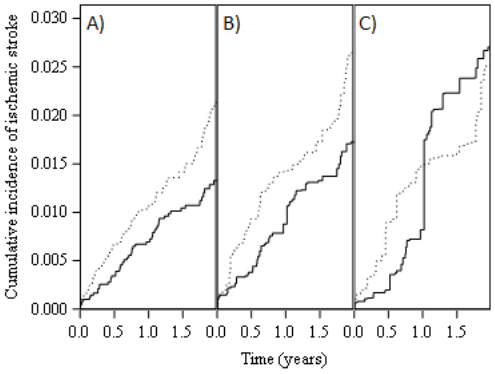

Figure 27: Cumulative incidence of all-cause mortality during two years of follow-up and censoring after discontinuation and applying censoring weights in A) the on-treatment trial population without sampling weights, B) the trial population weighted to the less-frail initiators of dabigatran in Medicare (P1), C) the trial population weighted to the less-frail initiators of warfarin in Medicare (P2), D) the trial population weighted to initiators of dabigatran in Medicare (P3), and E) the trial population weighted to initiators of warfarin in Medicare (P4). The solid line shows survival for dabigatran users and the dashed line corresponds to warfarin users.

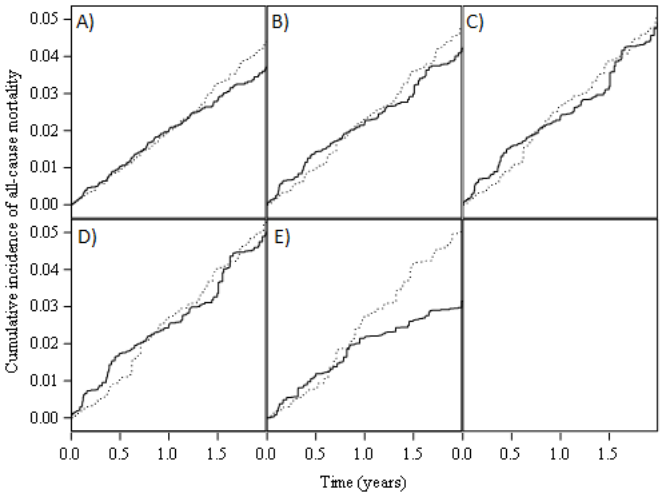

Figure 28: Cumulative incidence of gastrointestinal bleeding during two years of follow-up and censoring after discontinuation and applying censoring weights in A) the on-treatment trial population without sampling weights, B) the trial population weighted to the less-frail initiators of dabigatran in Medicare (P1), C) the trial population weighted to the less-frail initiators of warfarin in Medicare (P2), D) the trial population weighted to initiators of dabigatran in Medicare (P3), and E) the trial population weighted to initiators of warfarin in Medicare (P4). The solid line shows survival for dabigatran users and the dashed line corresponds to warfarin users.

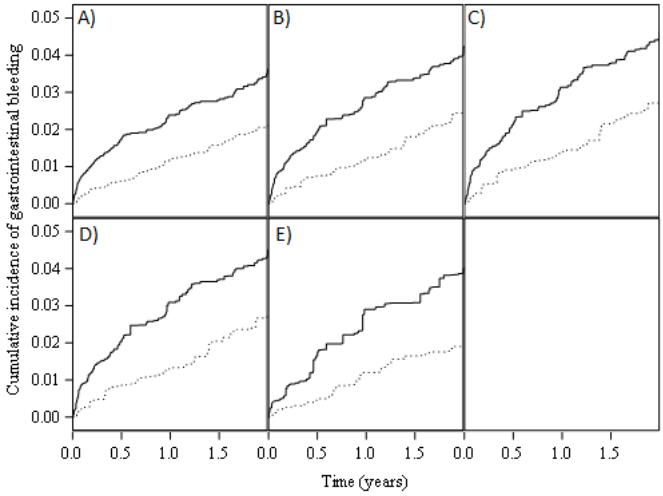

Figure 29: Cumulative incidence of all stroke during two years of follow-up and censoring after discontinuation and applying censoring weights in A) the on-treatment trial population without sampling weights, B) the trial population weighted to the less-frail initiators of dabigatran in Medicare (P1), C) the trial population weighted to the less-frail initiators of warfarin in Medicare (P2), D) the trial population weighted to initiators of dabigatran in Medicare (P3), and E) the trial population weighted to initiators of warfarin in Medicare (P4). The solid line shows survival for dabigatran users and the dashed line corresponds to warfarin users.

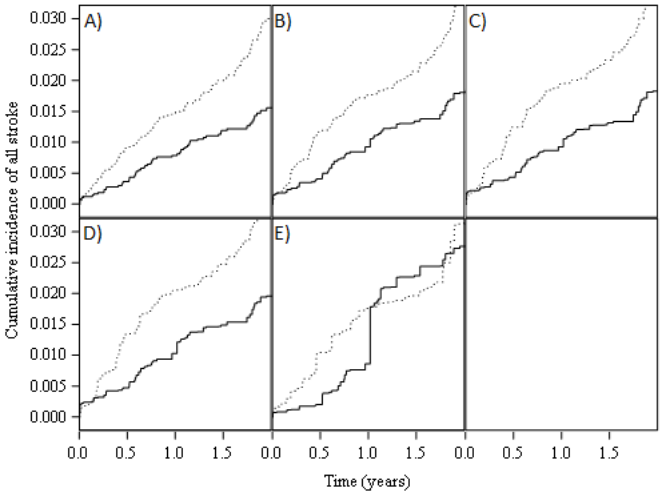

Figure 30: Cumulative incidence of major bleeding during two years of follow-up and censoring after discontinuation and applying censoring weights in A) the on-treatment trial population without sampling weights, B) the trial population weighted to the less-frail initiators of dabigatran in Medicare (P1), C) the trial population weighted to the less-frail initiators of warfarin in Medicare (P2), D) the trial population weighted to initiators of dabigatran in Medicare (P3), and E) the trial population weighted to initiators of warfarin in Medicare (P4). The solid line shows survival for dabigatran users and the dashed line corresponds to warfarin users.

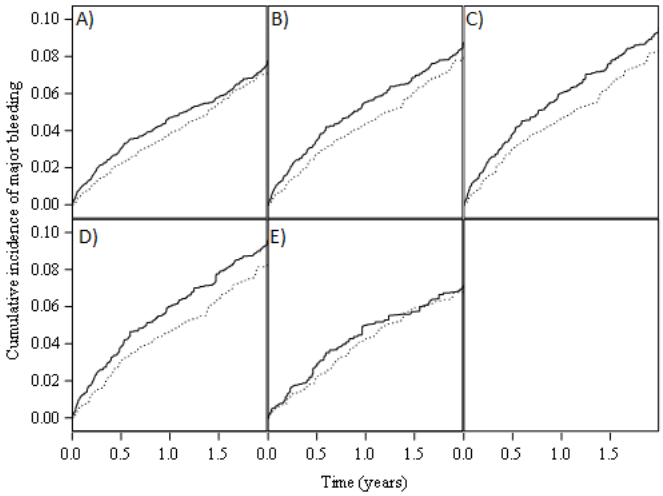

Table 13: Person-years, event numbers, two-year risks, and two-year on-treatment risk differences for ischemic stroke, mortality, and gastrointestinal bleeding in RE-LY trial participants over 65 after reweighting[a] to resemble dabigatran initiators (P1) or warfarin initiators (P2) in Medicare with less than 15% predicted probability of frailty.

Outcome and target population	Person-years	Events	Two-year risk	Two-year risk difference RD_{TP} (95% CI)
Ischemic stroke-P1				
Warfarin	7,951	1,457	2.40%	Ref.
Dabigatran 150 mg	7,508	1,240	1.62%	-0.78% (-1.69%, 0.14%)
Ischemic stroke-P2				
Warfarin	7,911	1,522	2.54%	Ref.
Dabigatran 150 mg	7,489	1,389	1.63%	-0.91% (-1.93%, 0.11%)
All-cause mortality-P1				
Warfarin	8,003	630	4.79%	Ref.
Dabigatran 150 mg	7,535	544	4.22%	-0.57% (-1.83%, 0.68%)
All-cause mortality-P2				
Warfarin	7,967	661	5.07%	Ref.
Dabigatran 150 mg	7,518	625	4.90%	-0.16% (-1.71%, 1.39%)
Gastrointestinal bleeding-P1				
Warfarin	7,944	1,469	2.48%	Ref.
Dabigatran 150 mg	7,435	1,479	4.23%	1.75% (0.76%, 2.74%)
Gastrointestinal bleeding-P2				
Warfarin	7,900	1,546	2.80%	Ref.
Dabigatran 150 mg	7,391	1,689	4.65%	1.85% (0.65%, 3.04%)

CI = confidence interval. P1= the population of less-frail dabigatran initiators. P2=the population of less-frail warfarin initiators.

[a]Weighted with inverse odds of sampling weights including age modeled with restricted cubic splines at the 20th, 40th, 60th, and 80th percentile, sex, use of amiodarone or proton pump inhibitors, current smoking, diabetes, congestive heart failure, coronary artery disease, past stroke, past transient ischemic attack, and past cancer.

Table 14: Person-years, event numbers, two-year risks, and two-year on-treatment risk differences for all stroke and major bleeding in RE-LY trial participants over 65 after reweighting[a] to resemble dabigatran initiators (P1) or warfarin initiators (P2) in Medicare.

Outcome and target population	Person-years	Events	Two-year risk	Two-year risk difference RD_{TP} (95% CI)
All stroke-P1				
Warfarin	7,941	1,468	3.32%	
Dabigatran 150 mg	7,506	1,232	1.81%	-1.52% (-2.52%, -0.51%)
All stroke-P2				
Warfarin	7,898	1,533	3.53%	
Dabigatran 150 mg	7,488	1,377	1.83%	-1.70% (-2.80%, -0.59%)
Major bleeding-P1				
Warfarin	7,780	2,054	7.90%	
Dabigatran 150 mg	7,329	1,910	8.73%	0.83% (-0.63%, 2.29%)
Major bleeding-P2				
Warfarin	7,726	2,145	8.38%	
Dabigatran 150 mg	7,270	2,126	9.59%	1.2% (-0.45%, 2.86%)

CI = confidence interval. P1= the population of less-frail dabigatran initiators. P2=the population of less-frail warfarin initiators.

[a]Weighted with inverse odds of sampling weights including age modeled with restricted cubic splines at the 20th, 40th, 60th, and 80th percentile, sex, use of amiodarone or proton pump inhibitors, current smoking, diabetes, congestive heart failure, coronary artery disease, past stroke, past transient ischemic attack, and past cancer.

Table 15: Person-years, event numbers, two-year risks, and two-year as-treated risk differences for ischemic stroke, mortality, and gastrointestinal bleeding in RE-LY trial participants over 65 after reweighting to resemble dabigatran initiators (P3) or warfarin initiators (P4) in Medicare.

Outcome and target population	**Person-years**	**Events**	**Two-year risk**	**Two-year risk difference RD_{TP} (95% CI)**
Ischemic stroke-P3				
Warfarin	7,772	1,556	2.65%	Ref.
Dabigatran 150 mg	7,462	1,437	1.72%	-0.92% (-2.02%, 0.17%)
Ischemic stroke-P4				
Warfarin	6,476	1,301	2.52%	Ref.
Dabigatran 150 mg	5,827	842	2.7%	0.18% (-2.05%, 2.41%)
All-Cause mortality-P3				
Warfarin	7,832	671	5.27%	Ref.
Dabigatran 150 mg	7,497	644	5.01%	-0.26% (-1.92%, 1.39%)
All-Cause mortality-P4				
Warfarin	6,530	574	5.02%	Ref.
Dabigatran 150 mg	5,858	292	3.15%	-1.87% (-3.36%, -0.37%)
Gastrointestinal bleeding-P3				
Warfarin	7,769	1,566	2.73%	Ref.
Dabigatran 150 mg	7,375	1,701	4.51%	1.78% (0.6%, 2.96%)
Gastrointestinal bleeding-P4				
Warfarin	6,485	1,304	1.91%	Ref.
Dabigatran 150 mg	5,788	924	4.01%	2.1% (0.51%, 3.69%)
All stroke-P3				
Warfarin	7,758	1,569	3.67%	Ref.
Dabigatran 150 mg	7,460	1,427	1.96%	-1.71% (-2.92%, -0.51%)
All stroke-P4				
Warfarin	6,468	1,356	3.14%	Ref.
Dabigatran 150 mg	5,826	844	2.77%	-0.37% (-2.62%, 1.88%)
Major bleeding-P3				
Warfarin	7,592	2,162	8.31%	Ref.
Dabigatran 150 mg	7,245	2,172	9.58%	1.26% (-0.53%, 3.06%)
Major bleeding-P4				
Warfarin	6,343	1,734	6.79%	Ref.
Dabigatran 150 mg	5,721	1,198	7.12%	0.34% (-1.78%, 2.45%)

CI = confidence interval. P3= the population of all dabigatran initiators. P4=the population of warfarin initiators.
[a]Weighted with inverse odds of sampling weights including age modeled with restricted cubic splines at the 20th, 40th, 60th, and 80th percentile, sex, use of amiodarone or proton pump inhibitors, current smoking, diabetes, congestive heart failure, coronary artery disease, past stroke, past transient ischemic attack, and past cancer.

Table 16: Two-year risk differences for the three main outcomes obtained using SMR weighting in all four main target populations.

Outcome	SMR two-year RD in less-frail dabigatran new users, RD_{O1} (95% CI)	SMR two-year RD in less-frail warfarin new users, RD_{O2} (95% CI)	SMR two-year RD in dabigatran new users, RD_{O3} (95% CI)	SMR two-year RD in warfarin new users, RD_{O1} (95% CI)
Ischemic Stroke	-0.46% (-0.88%, -0.03%)	-0.55% (-1.22%, 0.13%)	-0.67% (-1.10%, -0.24%)	-0.74% (-1.47%, -0.01%)
Mortality	-1.94% (-2.92%, -0.97%)	-0.39% (-2.68%, 2.09%)	-2.96% (-3.97%, -1.95%)	-3.00% (-5.23%, -0.74%)
Gastrointestinal bleeding	0.14% (-0.82%, 1.09%)	0.57% (-0.70%, 1.83%)	0.51% (-0.30%, 1.31%)	1.97% (-0.15%, 4.08%)

CI = Confidence interval. SMR = standardized morbidity ratio. RD = risk difference. RD_{OP} = the SMR-weighted risk difference using non-experimental data in target population P.

[a]Weighted based upon logistic regression including frailty and age modeled with restricted cubic splines at the 20th, 40th, 60th, and 80th percentile, sex, diabetes, coronary artery disease, congestive heart failure, hypertension, stroke in the past year, transient ischemic attack, cancer, bleed and GI bleed history, past use of vitamin K antagonists, alcohol abuse, acute renal problems, atherosclerosis, cardioversion in the past year, deep vein thrombosis in the past year, hyperlipidemia, obesity, pulmonary embolism in the past year, peripheral vascular disease and codes indicating smoking.

Figure 31: Each panel plots the estimated risk differences from the transported RE-LY trial on the vertical axis against the estimated risk differences from non-experimental methods in those same populations on the horizontal axis, with each panel corresponding to a different outcome (panel A is ischemic stroke; panel B is mortality; and panel C is gastrointestinal bleeding) and each point representing a different target population (circles are the target population of less-frail dabigatran users, squares are the target population of less-frail warfarin users). The closer the point estimates are to the reference line at 45 degrees, the more similar the findings are. If the horizontal line crosses that line, 95% confidence limits of the non-experimental estimate include the weighted trial estimate; if the vertical line crosses that line, 95% confidence limits of the weighted trial estimate include the non-experimental estimate.

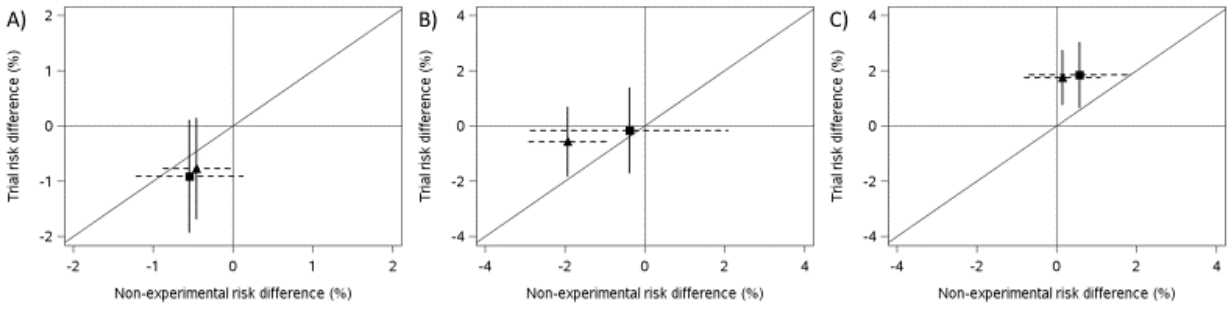

Figure 32: Each panel plots the estimated risk differences from the transported RE-LY trial on the vertical axis against the estimated risk differences from non-experimental methods in those same populations on the horizontal axis, with each panel corresponding to a different outcome (panel A is ischemic stroke; panel B is mortality; and panel C is gastrointestinal bleeding) and each point representing a different target population (circles are the target population of dabigatran users, squares are the target population of warfarin users). The closer the point estimates are to the reference line at 45 degrees, the more similar the two results are. If the horizontal line crosses that line, 95% confidence limits of the non-experimental estimate include the weighted trial estimate; if the vertical line crosses that line, 95% confidence limits of the weighted trial estimate include the non-experimental estimate.

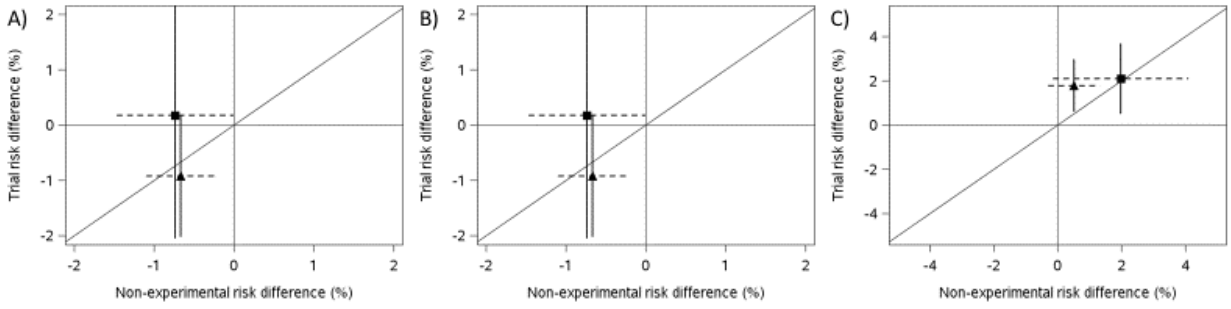

Table 17**:** Two-year risks and risk differences comparing ischemic stroke, mortality, gastrointestinal bleeding, all stroke, and major bleeding in arms of the weighted RELY trial to the observed risks in dabigatran (T1) and warfarin (T2) users in Medicare with less than 15% predicted probability of frailty,

Outcome and target population	**Two-year risk for the weighted RE-LY population on the corresponding[a] drug, R_T**	**Observed two-year risk in the Medicare population on the corresponding[a] drug, R_O**	**Two-year risk difference for R_T-R_O (95% CI)**
Ischemic stroke			
Dabigatran users (P1)	1.62%	1.05%	-0.57% (-1.20%, 0.05%)
Warfarin users (P2)	2.54%	1.85%	-0.69% (-1.48%, 0.09%)
All-cause mortality			
Dabigatran users (P1)	4.22%	4.4%	0.18% (-1.07%, 1.44%)
Warfarin users (P2)	5.07%	7.09%	2.02% (0.84%, 3.20%)
Gastrointestinal bleeding			
Dabigatran users (P1)	4.23%	3.33%	-0.90% (-2.05%, 0.25%)
Warfarin users (P2)	2.8%	3.16%	0.36% (-0.41%, 1.13%)
All stroke			
Dabigatran users (P1)	1.81%	1.38%	-0.43% (-1.18%, 0.32%)
Warfarin users (P2)	3.53%	2.22%	-1.31% (-2.16%, -0.45%)
Major bleeding			
Dabigatran users (P1)	8.73%	3.97%	-4.76% (-6.23%, -3.29%)
Warfarin users (P2)	8.38%	4.33%	-4.05% (-5.13%, -2.98%)

CI=confidence interval.

[a]For the target population P1 of dabigatran users, this risk is the risk in dabigatran users in the weighted RE-LY or Medicare population in the left and right column, respectively. For the population P2 of warfarin users, this is the risk in warfarin users in the weighted RE-LY or Medicare population.

Table 18: Two-year risks and risk differences comparing ischemic stroke, mortality, gastrointestinal bleeding, all stroke, and major bleeding in arms of the weighted RELY trial to the observed risks in all dabigatran and warfarin users in Medicare.

Outcome and target population	**Two-year risk for the weighted RE-LY population on the corresponding[a] drug, R_T**	**Observed two-year risk in the Medicare population on the corresponding[a] drug, R_O**	**Two-year risk difference for R_T-RO (95% CI)**
Ischemic stroke			
Dabigatran users (P3)	1.72%	1.29%	-0.43% (-1.12%, 0.26%)
Warfarin users (P4)	2.52%	2.48%	-0.04% (-1.12%, 1.04%)
All-cause mortality			
Dabigatran users (P3)	5.01%	6.82%	1.81% (0.16%, 3.45%)
Warfarin users (P4)	5.02%	14.02%	9.00% (7.71%, 10.29%)
Gastrointestinal bleeding			
Dabigatran users (P3)	4.51%	4%	-0.51% (-1.78%, 0.75%)
Warfarin users (P4)	1.91%	3.76%	1.85% (1.14%, 2.56%)
All stroke			
Dabigatran users (P3)	1.96%	1.59%	-0.37% (-1.15%, 0.41%)
Warfarin users (P4)	3.14%	2.99%	-0.15% (-1.29%, 0.99%)
Major bleeding			
Dabigatran users (P3)	9.58%	4.64%	-4.94% (-6.62%, -3.26%)
Warfarin users (P4)	6.79%	5.19%	-1.60% (-3.01%, -0.20%)

CI=confidence interval.

[a]For the population P3 of dabigatran users, this risk is the risk in dabigatran users in the weighted RE-LY or Medicare population in the left and right column, respectively. For the population P4 of warfarin users, this is the risk in warfarin users in the weighted RE-LY or Medicare population.

Table 19**:** Risk differences for various outcomes from sensitivity analyses after applying additional exclusion criteria and further frailty restrictions.

Outcome	Two-year RD (95% CI) after weighting to ≤10% PPF dabigatran new users	Two-year RD (95% CI) after weighting to ≤10% PPF warfarin new users	Two-year RD (95% CI) after weighting to dabigatran new users with no stroke in 180 days	Two-year RD (95% CI) after weighting to warfarin new users with no stroke in 180 days.
Ischemic stroke	-0.93% (-1.82%, -0.04%)	-0.72% (-1.67%, 0.24%)	-0.82% (-1.77%, 0.13%)	-0.19% (-1.30%, 0.91%)
Mortality	-0.64% (-1.80%, 0.51%)	-2.17% (-3.52%, -0.82%)	-0.28% (-1.67%, 1.10%)	-1.67% (-3.77%, 0.43%)
Gastrointestinal bleeding	1.71% (0.77%, 2.66%)	0.72% (-0.37%, 1.81%)	1.63% (0.51%, 2.76%)	1.96% (0.30%, 3.62%)
All stroke	-1.63% (-2.60%, -0.67%)	-1.36% (-2.4%, -0.33%)	-1.47% (-2.48%, -0.46%)	-0.66% (-1.83%, 0.50%)
Major bleeding	0.89% (-0.50%, 2.28%)	-0.76% (-2.28%, 0.77%)	0.96% (-0.61%, 2.53%)	0.30% (-1.83%, 2.44%)

PPF = predicted probability of frailty. CI = confidence interval.

Table 20**:** Risk differences for various outcomes from sensitivity analyses including altered hypertension and past vitamin K usage or a quadratic term for age in the sampling model.

Outcome	Two-year RD (95% CI) after weighting to ≤15% PPF dabigatran new users with additional covariates	Two-year RD (95% CI) after weighting to ≤15% PPF warfarin new users with additional covariates	Two-year RD (95% CI) after weighting to ≤15% PPF dabigatran new users with quadratic age term	Two-year RD (95% CI) after weighting to ≤15% PPF warfarin new users with quadratic age term
Ischemic stroke	-0.33% (-1.90%, 1.23%)	-0.63% (-2.15%, 0.90%)	-0.82% (-1.70%, 0.05%)	-0.97% (-1.94%, -0.01%)
Mortality	-0.73% (-2.62%, 1.15%)	-0.60% (-2.76%, 1.55%)	-0.53% (-1.73%, 0.66%)	-0.13% (-1.59%, 1.32%)
Gastrointestinal bleeding	1.60% (-0.03%, 3.23%)	1.46% (-0.19%, 3.10%)	1.82% (0.88%, 2.76%)	1.97% (0.84%, 3.10%)
All stroke	-1.03% (-2.68%, 0.63%)	-1.38% (-2.96%, 0.20%)	-1.56% (-2.53%, -0.59%)	-1.76% (-2.82%, -0.71%)
Major bleeding	0.44% (-1.96%, 2.83%)	0.78% (-1.41%, 2.96%)	0.91% (-0.54%, 2.36%)	1.28% (-0.33%, 2.90%)

PPF = predicted probability of frailty. CI = confidence interval.

CHAPTER 6: CONCLUSIONS

6.1: Main findings

We had two main goals in this work: first, estimate treatment effects of dabigatran vs warfarin on a variety of outcomes with propensity score methods and a new user active comparator design entirely using non-experimental data; and second, estimate analogous treatment effects using inverse odds of sampling weights and data from a randomized trial combined with target population data. In the first goal, we also explored differences in effects of initiating and staying on treatment (adherence-adjusted or on-treatment design) versus initial treatment assignment (first treatment carried forward or intention-to-treat design).

In the first goal, we relied entirely on data from a 20% random sample of Medicare beneficiaries provided by the Centers for Medicare and Medicaid Services. We used outcome and covariate data from the dabigatran and warfarin initiators we were intending to use as target populations in the first goal to estimate both as-treated and intention-to-treat effects. Unlike the transported trial, which focused on estimating effects in warfarin and dabigatran initiators separately, the non-experimental work estimated treatment effects in the entire population (estimated using inverse probability of treatment weights, IPTW) as well as treatment effects in initiators of dabigatran and warfarin separately (estimated using standardized morbidity ratio, SMR, weights).

As-treated estimates of treatment effect in all initiators were similar to past non-experimental work: dabigatran appeared protective for ischemic stroke (two-year IPTW RD, RD_{IPTW}: -0.73%, 95% CI: -1.40%, -0.06%), strongly protective for mortality (RD_{IPTW}: 2.98%, 95% CI: -5.05%, -0.91%), and harmful for gastrointestinal bleeding (RD_{IPTW}: 1.79%, 95% CI: -0.13%, 3.71%). Gastrointestinal bleeding harms appeared weaker in initiators of dabigatran (two-year SMR RD, RD_{dabi}: 0.51%, 95% CI: -0.30%, 1.31%), but effects on mortality (RD_{dabi}: -2.96%, 95% CI -3.97%, -1.95%) and ischemic stroke (RD_{dabi}: -0.67%, 95% CI: -1.10%, -0.24%) stayed relatively constant.

Intention-to-treat estimates, however, told a very different story. Dabigatran actually appeared to be harmful with respect to ischemic stroke in all users (RD_{IPTW}: 0.44%, 95% CI -0.22%, 1.09%), and mortality benefits were greatly attenuated (RD_{IPTW}: -0.84%, 95% CI: -2.39%, 0.72%). Gastrointestinal bleeding effects were also attenuated (RD_{IPTW}: 1.05%, 95% CI: 0.08%, 2.01%). Targeting dabigatran initiators reduced, but did not eliminate, the change in estimates for ischemic stroke (RD_{dabi}: 0.16%, 95% CI -0.20%, 0.52%), mortality (RD_{dabi}: -1.65%, 95% CI -2.32%, -0.98%) and gastrointestinal bleeding (RD_{dabi}: 0.36%, 95% CI -0.08%, 0.79%). These differences in intention-to-treat and as-treated estimates remained across a variety of sensitivity analyses. While it's theoretically possible that this could be due to an open backdoor path through factors predicting treatment adherence biasing intention-to-treat estimates but not as-treated estimates, it does emphasize the potential for loss of clinically relevant information when we only examine effects in those that remain on treatment.

Meanwhile, to accomplish the second goal, we leaned heavily on both the Randomized Evaluation of Long-Term Anticoagulation (RE-LY) trial and the 20% random sample of Medicare beneficiaries. After reweighting the trial participants to resemble two target populations of dabigatran initiators and warfarin initiators in the target population and censoring at treatment discontinuation, there was evidence of issues with positivity towards the edges of the the sampling distribution and inability to match mortality in the target population.

After restricting ourselves to Medicare initiators with less than 15% predicted probability of frailty to mimic trial exclusion criteria, however, these issues diminished, and we were able to obtain as-treated estimates of treatment effect for ischemic stroke (RD_{dabi}: -0.77%, 95% CI -1.69%, 0.14%; RD_{warf}: -0.91%, 95% CI -1.93%, 0.11%), all-cause mortality (RD_{dabi}: -0.57%, 95% CI -1.83%, 0.68%; RD_{warf}: -0.16%, 95% CI -1.71%, 1.39%), and gastrointestinal bleeding (RD_{dabi}: 1.75%, 95% CI 0.76%, 2.74%; RD_{warf}: 1.85%, 95% CI 0.65%, 3.04%). Even in this restricted population, however, differences remained in mortality for warfarin users relative to Medicare initiators, and after applying those weights we saw more stroke and gastrointestinal bleeding events in weighted trial patients than observed in our targets. Differences in warfarin therapy, differences in outcome sensitivity and specificity, and missing information on the key effect measure modifier of bleed were all potential causes of these differences.

To have appropriate benchmarks for comparing the non-experimental estimate to the transported estimates, we also re-analyzed the non-experimental data limiting ourselves to initiators with less than 15% predicted probability of frailty and specifically targeting the two populations of dabigatran and warfarin users. These SMR weighted estimates in the non-experimental study showed very similar estimates of effect on ischemic stroke risk to the reweghted trial (RD_{dabi}: -0.46%, 95% CI -0.88%, -0.03%; RD_{warf}: -0.55%, 95% CI -1.22%, 0.13%), but estimates of effects on mortality were larger for dabigatran users (RD_{dabi}: -1.94%, 95% CI -2.92%, -0.97%; RD_{warf}: -0.39%, 95% CI -2.68%, 2.09%) and estimates of effects on gastrointestinal bleeding were reduced (RD_{dabi}: 0.14%, 95% CI -0.82%, 1.09%; RD_{warf}: 0.57%, 95% CI -0.70%, 1.83%). These differences in mortality and gastrointestinal bleeding estimates could be due to unmeasured confounding, measurement error, differences in distributions of unmeasured effect measure modifiers, differences in warfarin management, or all three.

6.2: Significance

With the ongoing phasing out of warfarin in favor of dabigatran and other novel oral anticoagulants in patients with atrial fibrillation, understanding the advantages and disadvantages of each of these new drugs is key. The differences in intention-to-treat and as-treated estimates from the non-experimental work reemphasize the importance of remaining on a stabilized treatment for patients receiving oral anticoagulants. While it is tempting to limit ourselves to estimating one of these two intervention effects, the context gained from estimating both is vital to simulaneously improve clinician decision-making and public health.

This work is provides the first absolute-scale effect estimates from RE-LY including censoring weights and some of the first absolute-scale effect estimates in Medicare as whole. Additionally, this is the first time the trial data has been reweighted to take into account differing distributions of risk factors in older adults receiving routine care for atrial fibrillation. Differences in effects on gastrointestinal bleeding and mortality in even the frailty-restricted trial and non-experimental study lends significant context the large mortality benefits from this and other non-experimental work, though it is possible those benefits are concentrated heavily in the frailer patients and based on differing patterns of warfarin therapy.

Finally, we also observed potential differences in treatment effect across target populations of warfarin users and dabigatran users, suggesting that prescribers may have been successful in steering individuals towards dabigatran for whom it is a comparatively better option for some outcomes.

6.3: Future directions

We plan to use this work as a case study illustrating the potential impact of misclassification and missing data for transporting trial results to target populations. This could help contextualize the extent to which problems we observed in the differential classification of effect measure modifiers we wanted to include in our sampling models (e.g. hypertension, history of major bleeding, and past vitamin K antagonist use) could bias results in this and future studies. We also hope to use this scenario to assess potential exaggeration of chance confounding after randomization by effect modifiers. We further hope to contrast transported effects in all dabigatran new users with transported effects in individuals switching to dabigatran. Finally, examining per-protocol treatment effects with a more complex treatment rule, rather than simply requiring individuals stay on treatment, could provide additional insight into treatment effects, as could digging further into the components of the censoring models and attempting to understand the differential experience of warfarin and dabigatran switchers and stoppers.

We strongly suggest that future non-experimental pharmacoepidemiologic studies assessing long-term treatments consider estimating both intention-to-treat and as-treated treatment effects. Determining and providing in publications which drug had longer follow-up time in the as-treated approach is similarly vital. This is especially important in an active comparator new user design where one drug may have radically differing adherence and persistence patterns relative to the other. Knowing when differential stopping or switching rates cause overall worse performance is key to public health.

As a result of this, additional work in understanding the best ways to transport intention-to-treat effect estimates is key. Can trial populations with much lower rates of discontinuation and re-initiation be used to stand-in for target populations that stop, switch, and restart treatment at double the rate? Can and should this be done routinely for trials with run-in periods or restrictions designed to improve adherence? Or should the focus be on estimating treatment effects that follow specific protocols?

We recommend future work transporting study results to other populations where outcome data is available should make sure to assess whether the weighted populations experience similar outcomes to those from the original to identify when key covariates may be unavailable for the sampling model. Researchers should also be sure to assess how covariates in the source and target populations have been measured differently when determining whether the fits of their sampling models are plausible. Additional work on methods for identifying whether covariates are necessary for a sampling model would also help precision.

Finally, we suggest comparisons of non-experimental and trial data in both this and other settings should do their best to move beyond direct comparisons of point estimates and assuming there is no heterogeneity of treatment effect, particularly when there may be issues of non-positivity in the trial. When non-experimental data and trial data conflict, it's vital to address as many potential sources of error as possible, including standardizing both target population and intention-to-treat versus as-treated estimands. Once that's done, considering the plausibility of confounding, measurement error, and missing effect measure modifiers is paramount to making inferences about treatment effects for public health.

APPENDIX A: OUTCOME CODING

Outcome	Hospital Discharge Codes
Ischemic stroke	As primary discharge diagnosis: 433.x1 Occlusion and stenosis of precerebral arteries 434.x1 Occlusion and stenosis of cerebral arteries with cerebral infarction
All stroke	As primary discharge diagnosis: 431.x Intracerebal hemorrhage 433.x1 Occlusion and stenosis of precerebral arteries 434.x1 Occlusion and stenosis of cerebral arteries with cerebral infarction 436.x Acute but ill-defined cerebrovascular events
Gastrointestinal bleeding:	As primary discharge diagnosis: *Upper gastrointestinal bleed:* 531.0x Acute gastric ulcer with hemorrhage with or without obstruction 531.2x Acute gasric ulcer with hemorrhage and perforation with or without obstruction 531.4x Chronic or unspecified gastric ulcer with hemorrhage with or without obstruction 531.6x Gastric ulcer with hemorrhage and perforation with or without obstruction 532.0x Acute duodenal ulcer with hemorrhage with or without obstruction 532.2x Acute duodenal ulcer with hemorrhage and perforation with or without obstruction 532.4x Chronic or unspecified duodenal ulcer with hemorrhage with or without obstruction 532.6x Chronic or unspecified duodenal ulcer with hemorrhage and perforation with or without obstruction 533.0x Acute peptic ulcer of unspecified site with hemorrhage with or without obstruction 533.2x Acute peptic ulcer of unspecified site with hemorrhage and perforation with or without obstruction 533.4x Chronic or unspecified peptic ulcer of unspecified site with hemorrhage with or without obstruction 533.6x Chronic or unspecified peptic ulcer of unspecified site with hemorrhage and perforation with or without obstruction 534.0x Acute gastrojejunal ulcer with hemorrhage with or without obstruction

	534.2x Acute gastrojejunal ulcer with hemorrhage and perforation with or without obstruction 534.4x Chronic or unspecified gastrojejunal ulcer with hemorrhage with or without obstruction 534.6x Chronic or unspecified gastrojejunal ulcer with hemorrhage and perforation with or without obstruction 578.0 Hematemesis ICD-9 procedure code 44.43 Endoscopic control of gastric or duodenal bleeding CPT code 43255 Upper gastrointestinal endoscopy including esophagus, stomach and either the duodenum and/or jejunum as appropriate with control of bleeding, any method *Lower and unspecified G.I. bleeds:* 562.02 Diverticulosis of small intestine with hemorrhage 562.03 Diverticulitis of small intestine with hemorrhage 562.12 Diverticulosis of colon with hemorrhage 562.13 Diverticulitis of colon with hemorrhage 569.3x Hemorrhage of rectum and anus 569.85 Angiodysplasia of intestine with hemorrhage 578.1x Blood in stool 578.9 Hemorrhage of GI tract, unspecified
Major bleeding	As primary discharge diagnosis: *Intracranial bleeding:* 430.x Subarachnoid hemorrhage 431.x Intracerebral hemorrhage 432.x Other and unspecified intracranial hemorrhage *Upper gastrointestinal bleed:* 531.0x Acute gastric ulcer with hemorrhage with or without obstruction 531.2x Acute gasric ulcer with hemorrhage and perforation with or without obstruction 531.4x Chronic or unspecified gastric ulcer with hemorrhage with or without obstruction 531.6x Gastric ulcer with hemorrhage and perforation with or without obstruction 532.0x Acute duodenal ulcer with hemorrhage with or without obstruction

	532.2x Acute duodenal ulcer with hemorrhage and perforation with or without obstruction 532.4x Chronic or unspecified duodenal ulcer with hemorrhage with or without obstruction 532.6x Chronic or unspecified duodenal ulcer with hemorrhage and perforation with or without obstruction 533.0x Acute peptic ulcer of unspecified site with hemorrhage with or without obstruction 533.2x Acute peptic ulcer of unspecified site with hemorrhage and perforation with or without obstruction 533.4x Chronic or unspecified peptic ulcer of unspecified site with hemorrhage with or without obstruction 533.6x Chronic or unspecified peptic ulcer of unspecified site with hemorrhage and perforation with or without obstruction 534.0x Acute gastrojejunal ulcer with hemorrhage with or without obstruction 534.2x Acute gastrojejunal ulcer with hemorrhage and perforation with or without obstruction 534.4x Chronic or unspecified gastrojejunal ulcer with hemorrhage with or without obstruction 534.6x Chronic or unspecified gastrojejunal ulcer with hemorrhage and perforation with or without obstruction 578.0 Hematemesis ICD-9 procedure code 44.43 Endoscopic control of gastric or duodenal bleeding CPT code 43255 Upper gastrointestinal endoscopy including esophagus, stomach and either the duodenum and/or jejunum as appropriate with control of bleeding, any method <u>*Lower and unspecified G.I. bleeds:*</u> 562.02 Diverticulosis of small intestine with hemorrhage 562.03 Diverticulitis of small intestine with hemorrhage 562.12 Diverticulosis of colon with hemorrhage 562.13 Diverticulitis of colon with hemorrhage 569.3x Hemorrhage of rectum and anus 569.85 Angiodysplasia of intestine with hemorrhage 578.1x Blood in stool 578.9 Hemorrhage of GI tract, unspecified

	Other major bleeds: 285.1x Acute posthemorrhagic anemia 423.0x Hemopericardium 459.0x Hemorrhage not specified 599.7 Hematuria 719.1x Hemathrosis 786.3x Hemoptysis 984.7x Epistaxis

APPENDIX B: ELIGIBILITY CRITERIA AND CODING

Variable	**Related ICD-9 codes**
Inclusion:	
Atrial fibrillation	427.31
Congestive heart failure	1 inpatient or 2 outpatient ICD-9 codes: 428.x 398.91 402.01 402.11 402.91 404.01 404.11 404.91 404.03 404.13 404.93
Past stroke	ICD-9 codes: 431.x 433.x 434.x 436.x 437.1 438.x
Past TIA	ICD-9 code: 435.x
Diabetes	1 hospital discharge or 2 outpatient ICD-9 for DM 250.x OR Dispensing of metformin, sulfonylureas, insulin, or other direct antidiabetic agent
Hypertension	ICD-9 diagnosis codes: 401.x-405.x OR Dispensing of CCB, ACEI, ARB, BB, thiazide diuretic, or other direct antihypertensive agent
Coronary Artery Disease (CAD)	At least 1 ICD-9 code from: 410.x-414.x 429.2 V45.81

Exclusion:	
Valvular heart disease and heart valve replacement	ICD-9 diagnosis codes: 394.x 395.x 396.x 397.x 398.9x V42.2 V43.3 ICD-9 procedure codes: 35.1x CPT codes: 33660-33665 33400-33403 33405 33420-33468 33420-33430 33460 33463-33468 33475 33496 0257T 0258T 0259T 0262T
Active liver disease	ICD-9 diagnosis codes: 070.x 571.x-573.x 456.0-456.2x 155.0 155.1 155.2 576.8 ICD-9 procedure codes: 39.1 42.91
Cancer within the last 6 months	ICD-9 diagnosis codes: 140.x-208.x 230.x-239.x
Severe renal disease requiring dialysis	ICD-9 procedure codes: 39.95 54.98 56.0 V56.8 CPT codes:

	90935-90993 99512 99559
Chronic renal insufficiency	ICD-9 codes: 582.x-583.x 585.x-587.x
Active or subacute endocarditis	ICD-9 codes: 421.1
Predicted probability of frailty > 15%	Calculated from Faurot et al.[118]
Other covariates	
Systemic embolism	ICD-9 codes: 444.x
Deep vein thrombosis	451.x, 453.x
Pulmonary embolism	415.11, 415.12, 415.19
Hyperlipidemia	ICD-9 codes: 272.0 272.2 272.4 OR Statins or other antihyperlipidemic
Atherosclerosis	ICD-9 codes: 440.9 414.x 429.2
Peripheral vascular disease	1 inpatient or 2 outpatient claims with the following codes: ICD-9 codes: 440.20-440.24 440.29-440.32 440.3 443.9 ICD-9 procedure codes: 38.08 38.09 38.18 38.48 38.49 39.25 39.5 39.9 HCPCs: 35256, 35286, 35351, 35355, 35361, 35363, 35371, 35372, 35381, 35454, 35456, 35459, 35470, 35473, 35474, 35482, 35483, 35485,

	35492, 35493, 35495, 35521, 35533, 35541, 35546, 35548, 35549, 35551, 35556, 35558, 35563, 35565, 35566, 35571, 35621, 35623, 35641, 35646, 35647, 35650, 35651, 35654, 35656, 35661, 35663, 35666, 35671
Acute renal disease (within the past month)	ICD-9 codes: 580.0 580.4 580.8 580.9 581.0 581.1 581.2 581.3 581.8 581.9 584.6 584.7 584.8 584.9
Smoking	ICD-9 code: 305.1 649.0x 989.84
Obesity	ICD-9 code: 278.00
Alcoholism	ICD-9: 94.61-94.63 94.67-94.69 303.x 305.0x 291.x 357.5x 425.5x 571.1x 571.2x 571.3x

APPENDIX C: PROCEDURE CODES FOR ANTICOAGULATION MANAGEMENT

Current procedural terminology (CPT) codes:	99363, 99364, 85610, 3555F, G0248, G0249, G0250

REFERENCES

1. Rahman F, Kwan GF, Benjamin EJ. Global epidemiology of atrial fibrillation. *Nature reviews Cardiology.* 2014;11(11):639-654.

2. Savelieva I, Bajpai A, Camm AJ. Stroke in atrial fibrillation: update on pathophysiology, new antithrombotic therapies, and evolution of procedures and devices. *Annals of medicine.* 2007;39(5):371-391.

3. Marini C, De Santis F, Sacco S, et al. Contribution of atrial fibrillation to incidence and outcome of ischemic stroke: results from a population-based study. *Stroke.* 2005;36(6):1115-1119.

4. Rosand J, Eckman MH, Knudsen KA, Singer DE, Greenberg SM. The effect of warfarin and intensity of anticoagulation on outcome of intracerebral hemorrhage. *Arch Intern Med.* 2004;164(8):880-884.

5. Andrade AA, Li J, Radford MJ, Nilasena DS, Gage BF. Clinical Benefit of American College of Chest Physicians versus European Society of Cardiology Guidelines for Stroke Prophylaxis in Atrial Fibrillation. *Journal of general internal medicine.* 2015;30(6):777-782.

6. Connolly SJ, Ezekowitz MD, Yusuf S, et al. Dabigatran versus Warfarin in Patients with Atrial Fibrillation. *New England Journal of Medicine.* 2009;361(12):1139-1151.

7. Desmaele S, Steurbaut S, Cornu P, Brouns R, Dupont AG. Clinical trials with direct oral anticoagulants for stroke prevention in atrial fibrillation: how representative are they for real life patients? *European journal of clinical pharmacology.* 2016;72(9):1125-1134.

8. Eichler HG, Abadie E, Breckenridge A, et al. Bridging the efficacy-effectiveness gap: a regulator's perspective on addressing variability of drug response. *Nature reviews Drug discovery.* 2011;10(7):495-506.

9. Tanislav C, Milde S, Schwartzkopff S, Misselwitz B, Sieweke N, Kaps M. Baseline characteristics in stroke patients with atrial fibrillation: clinical trials versus clinical practice. *BMC research notes.* 2015;8:262.

10. Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. *Trials.* 2015;16.

11. Lip GY, Pan X, Kamble S, et al. Major bleeding risk among non-valvular atrial fibrillation patients initiated on apixaban, dabigatran, rivaroxaban or warfarin: a "real-world" observational study in the United States. *International journal of clinical practice.* 2016;70(9):752-763.

12. Lip GY, Keshishian A, Kamble S, et al. Real-world comparison of major bleeding risk among non-valvular atrial fibrillation patients initiated on apixaban, dabigatran, rivaroxaban, or warfarin. A propensity score matched analysis. *Thrombosis and haemostasis.* 2016;116(5):975-986.

13. Banerjee A, Lane DA, Torp-Pedersen C, Lip GY. Net clinical benefit of new oral anticoagulants (dabigatran, rivaroxaban, apixaban) versus no treatment in a 'real world' atrial fibrillation population: a modelling analysis based on a nationwide cohort study. *Thrombosis and haemostasis.* 2012;107(3):584-589.

14. Sturmer T, Rothman KJ, Avorn J, Glynn RJ. Treatment effects in the presence of unmeasured confounding: dealing with observations in the tails of the propensity score distribution--a simulation study. *Am J Epidemiol.* 2010;172(7):843-854.

15. Graham DJ, Reichman ME, Wernecke M, et al. Cardiovascular, bleeding, and mortality risks in elderly Medicare patients treated with dabigatran or warfarin for non-valvular atrial fibrillation. *Circulation.* 2014:CIRCULATIONAHA. 114.012061.

16. Hernandez I, Baik SH, Piñera A, Zhang Y. Risk of bleeding with dabigatran in atrial fibrillation. *JAMA internal medicine.* 2015;175(1):18-24.

17. Villines TC, Schnee J, Fraeman K, et al. A comparison of the safety and effectiveness of dabigatran and warfarin in non-valvular atrial fibrillation patients in a large healthcare system. *Thrombosis and haemostasis.* 2015;114(6):1290-1298.

18. Seeger JD, Bykov K, Bartels DB, Huybrechts K, Zint K, Schneeweiss S. Safety and effectiveness of dabigatran and warfarin in routine care of patients with atrial fibrillation. *Thrombosis and haemostasis.* 2015;114(6):1277-1289.

19. Larsen TB, Skjoth F, Nielsen PB, Kjaeldgaard JN, Lip GY. Comparative effectiveness and safety of non-vitamin K antagonist oral anticoagulants and warfarin in patients with atrial fibrillation: propensity weighted nationwide cohort study. *BMJ.* 2016;353:i3189.

20. Go AS, Singer DE, Toh S, et al. Outcomes of Dabigatran and Warfarin for Atrial Fibrillation in Contemporary Practice: A Retrospective Cohort Study. *Annals of Internal Medicine.* 2017;167(12):845-854.

21. Molteni M, Cimminiello C. Warfarin and atrial fibrillation: from ideal to real the warfarin affaire. *Thrombosis Jounral.* 2014;12(1):5.

22. January CT, Wann LS, Alpert JS, et al. 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines and the Heart Rhythm Society. *Circulation.* 2014;130(23):2071-2104.

23. Molteni M, Polo Friz H, Primitz L, Marano G, Boracchi P, Cimminiello C. The definition of valvular and non-valvular atrial fibrillation: results of a physicians' survey. *Europace.* 2014;16(12):1720-1725.

24. Roy D, Talajic M, Nattel S, et al. Rhythm Control versus Rate Control for Atrial Fibrillation and Heart Failure. *New England Journal of Medicine.* 2008;358(25):2667-2677.

25. Watson T, Shantsila E, Lip GY. Mechanisms of thrombogenesis in atrial fibrillation: Virchow's triad revisited. *The Lancet.* 2009;373(9658):155-166.

26. Aguero-Torres H, Fratiglioni L, Guo Z, Viitanen M, Winblad B. Mortality from dementia in advanced age: a 5-year follow-up study of incident dementia cases. *Journal of clinical epidemiology.* 1999;52(8):737-743.

27. Frost L, Hune LJ, Vestergaard P. Overweight and obesity as risk factors for atrial fibrillation or flutter: the Danish Diet, Cancer, and Health Study. *The American journal of medicine.* 2005;118(5):489-495.

28. Kalra L, Lip GYH. Antithrombotic treatment in atrial fibrillation. *Heart (British Cardiac Society).* 2007;93(1):39-44.

29. Krahn AD, Manfreda J, Tate RB, Mathewson FA, Cuddy TE. The natural history of atrial fibrillation: incidence, risk factors, and prognosis in the Manitoba Follow-Up Study. *The American journal of medicine.* 1995;98(5):476-484.

30. Morillo CA, Banerjee A, Perel P, Wood D, Jouven X. Atrial fibrillation: the current epidemic. *Journal of Geriatric Cardiology: JGC.* 2017;14(3):195.

31. Lane DA, Skjøth F, Lip GY, Larsen TB, Kotecha D. Temporal Trends in Incidence, Prevalence, and Mortality of Atrial Fibrillation in Primary Care. *Journal of the American Heart Association.* 2017;6(5):e005155.

32. Miyasaka Y, Barnes ME, Gersh BJ, et al. Secular trends in incidence of atrial fibrillation in Olmsted County, Minnesota, 1980 to 2000, and implications on the projections for future prevalence. *Circulation.* 2006;114(2):119-125.

33. Schnabel RB, Yin X, Gona P, et al. 50 year trends in atrial fibrillation prevalence, incidence, risk factors, and mortality in the Framingham Heart Study: a cohort study. *The Lancet.* 2015;386(9989):154-162.

34. Turakhia MP, Shafrin J, Bognar K, et al. Economic burden of undiagnosed nonvalvular atrial fibrillation in the United States. *The American journal of cardiology.* 2015;116(5):733-739.

35. Reynolds MR, Essebag V. Economic burden of atrial fibrillation: implications for intervention. *Am J Pharm Benefits.* 2012;4(2):58-65.

36. Shehab A, Elnour AA, Bhagavathula AS, et al. Novel oral anticoagulants and the 73rd anniversary of historical warfarin. *Journal of the Saudi Heart Association.* 2016;28(1):31-45.

37. Kuruvilla M, Gurk-Turner C. A review of warfarin dosing and monitoring. *Proceedings (Baylor University Medical Center).* 2001;14(3):305-306.

38. Holbrook AM, Pereira JA, Labiris R, et al. Systematic overview of warfarin and its drug and food interactions. *Arch Intern Med.* 2005;165(10):1095-1106.

39. Kim MM, Metlay J, Cohen A, et al. Hospitalization costs associated with warfarin-related bleeding events among older community-dwelling adults. *Pharmacoepidemiology and Drug Safety.* 2010;19(7):731-736.

40. Pavani A, Naushad SM, Uma A, Kutala VK. Methodological issues in the development of a pharmacogenomic algorithm for warfarin dosing: comparison of two regression approaches. *Pharmacogenomics.* 2014;15(8):1125-1132.

41. Pavani A, Naushad SM, Rupasree Y, et al. Optimization of warfarin dose by population-specific pharmacogenomic algorithm. *The pharmacogenomics journal.* 2012;12(4):306-311.

42. French B, Wang L, Gage BF, Horenstein RB, Limdi NA, Kimmel SE. A systematic analysis and comparison of warfarin initiation strategies. *Pharmacogenetics and genomics.* 2016;26(10):445-452.

43. Lip GY, Lane DA. Stroke prevention in atrial fibrillation: a systematic review. *Jama.* 2015;313(19):1950-1962.

44. Granger CB, Alexander JH, McMurray JJ, et al. Apixaban versus warfarin in patients with atrial fibrillation. *The New England journal of medicine.* 2011;365(11):981-992.

45. Connolly SJ, Ezekowitz MD, Yusuf S, et al. Dabigatran versus warfarin in patients with atrial fibrillation. *The New England journal of medicine.* 2009;361(12):1139-1151.

46. Giugliano RP, Ruff CT, Braunwald E, et al. Edoxaban versus warfarin in patients with atrial fibrillation. *The New England journal of medicine.* 2013;369(22):2093-2104.

47. Patel MR, Mahaffey KW, Garg J, et al. Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. *The New England journal of medicine.* 2011;365(10):883-891.

48. Olesen JB, Sørensen R, Hansen ML, et al. Non-vitamin K antagonist oral anticoagulation agents in anticoagulant naïve atrial fibrillation patients: Danish nationwide descriptive data 2011–2013. *EP Europace.* 2014;17(2):187-193.

49. Kirley K, Qato DM, Kornfield R, Stafford RS, Alexander GC. NATIONAL TRENDS IN ORAL ANTICOAGULANT USE IN THE UNITED STATES, 2007–2011. *Circulation Cardiovascular Quality and Outcomes.* 2012;5(5):615-621.

50. Wessler JD, Grip LT, Mendell J, Giugliano RP. The P-glycoprotein transport system and cardiovascular drugs. *Journal of the American College of Cardiology.* 2013;61(25):2495-2502.

51. Schiele F, van Ryn J, Newsome C, et al. A specific antidote for dabigatran: functional and structural characterization. *Blood.* 2013;121(18):3554-3562.

52. Grottke O, van Ryn J, Spronk HM, Rossaint R. Prothrombin complex concentrates and a specific antidote to dabigatran are effective ex-vivo in reversing the effects of dabigatran in an anticoagulation/liver trauma experimental model. *Critical Care.* 2014;18(1):R27.

53. Pollack Jr CV, Reilly PA, Eikelboom J, et al. Idarucizumab for dabigatran reversal. *New England Journal of Medicine.* 2015;373(6):511-520.

54. da Silva R. Novel oral Anticoagulants in Non-Valvular Atrial Fibrillation. *Cardiovascular & Hematological Agents in Medicinal Chemistry.* 2014;12(1):3-8.

55. Connolly SJ, Milling TJ, Jr., Eikelboom JW, et al. Andexanet Alfa for Acute Major Bleeding Associated with Factor Xa Inhibitors. *New England Journal of Medicine.* 2016;375(12):1131-1141.

56. Sleight P. Debate: Subgroup analyses in clinical trials: fun to look at - but don't believe them! *Current Controlled Trials in Cardiovascular Medicine.* 2000;1(1):25-27.

57. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Ann Intern Med.* 2010;152(11):726-732.

58. Rothman KJ GS, Lash TL, editors. *Modern Epidemiology.* 3rd ed. New York, NY: Lippincott-Raven; 2008.

59. Eikelboom JW, Wallentin L, Connolly SJ, et al. Risk of bleeding with 2 doses of dabigatran compared with warfarin in older and younger patients with atrial fibrillation: an analysis of the randomized evaluation of long-term anticoagulant therapy (RE-LY) trial. *Circulation.* 2011;123(21):2363-2372.

60. Lauw MN, Eikelboom JW, Coppens M, et al. Effects of dabigatran according to age in atrial fibrillation. *Heart (British Cardiac Society).* 2017;103(13):1015-1023.

61. Diener H-C, Connolly SJ, Ezekowitz MD, et al. Dabigatran compared with warfarin in patients with atrial fibrillation and previous transient ischaemic attack or stroke: a subgroup analysis of the RE-LY trial. *The Lancet Neurology.* 2010;9(12):1157-1163.

62. Dans AL, Connolly SJ, Wallentin L, et al. Concomitant use of antiplatelet therapy with dabigatran or warfarin in the Randomized Evaluation of Long-term Anticoagulation Therapy (RE-LY®) trial. *Circulation.* 2012:CIRCULATIONAHA. 112.115386.

63. Flaker G, Ezekowitz M, Yusuf S, et al. Efficacy and safety of dabigatran compared to warfarin in patients with paroxysmal, persistent, and permanent atrial fibrillation. *Journal of the American College of Cardiology.* 2012;59(9):854-855.

64. Ezekowitz MD, Reilly PA, Nehmiz G, et al. Dabigatran with or without concomitant aspirin compared with warfarin alone in patients with nonvalvular atrial fibrillation (PETRO Study). *The American journal of cardiology.* 2007;100(9):1419-1426.

65. Schulman S, Kearon C, Kakkar AK, et al. Dabigatran versus warfarin in the treatment of acute venous thromboembolism. *New England Journal of Medicine.* 2009;361(24):2342-2352.

66. Larsen TB, Rasmussen LH, Skjøth F, et al. Efficacy and safety of dabigatran etexilate and warfarin in “real-world” patients with atrial fibrillation: a prospective nationwide cohort study. *Journal of the American College of Cardiology.* 2013;61(22):2264-2273.

67. Curtis LH, Weiner MG, Boudreau DM, et al. Design considerations, architecture, and use of the Mini-Sentinel distributed data system. *Pharmacoepidemiology and Drug Safety.* 2012;21(S1):23-31.

68. Zhang Z. Survival analysis in the presence of competing risks. *Annals of Translational Medicine.* 2017;5(3).

69. Rothwell PM. Factors That Can Affect the External Validity of Randomised Controlled Trials. *PLoS Clinical Trials.* 2006;1(1):e9.

70. Van Spall HG, Wallentin L, Yusuf S, et al. Variation in warfarin dose adjustment practice is responsible for differences in the quality of anticoagulation control between centers and countries: an analysis of patients receiving warfarin in the randomized evaluation of long-term anticoagulation therapy (RE-LY) trial. *Circulation.* 2012;126(19):2309-2316.

71. Reiffel JA. Time in the Therapeutic Range for Patients Taking Warfarin in Clinical Trials. *Circulation.* 2017;135(16):1475-1477.

72. Wan Y, Heneghan C, Perera R, et al. Anticoagulation control and prediction of adverse events in patients with atrial fibrillation: a systematic review. *Circ Cardiovasc Qual Outcomes.* 2008;1(2):84-91.

73. Gottlieb LK, Salem-Schatz S. Anticoagulation in atrial fibrillation: does efficacy in clinical trials translate into effectiveness in practice? *Archives of Internal Medicine.* 1994;154(17):1945-1953.

74. Dlott JS, George RA, Huang X, et al. A national assessment of warfarin anticoagulation therapy for stroke prevention in atrial fibrillation. *Circulation.* 2014:CIRCULATIONAHA. 113.002601.

75. Baker WL, Cios DA, Sander SD, Coleman CI. Meta-analysis to assess the quality of warfarin control in atrial fibrillation patients in the United States. *Journal of Managed Care Pharmacy.* 2009;15(3):244-252.

76. Han SY, Palmeri ST, Broderick SH, et al. Quality of anticoagulation with warfarin in patients with nonvalvular atrial fibrillation in the community setting. *Journal of Electrocardiology.* 2013;46(1):45-50.

77. Gallego P, Roldán V, Marin F, et al. SAMe-TT 2 R 2 score, time in therapeutic range, and outcomes in anticoagulated patients with atrial fibrillation. *The American journal of medicine.* 2014;127(11):1083-1088.

78. Wallentin L, Yusuf S, Ezekowitz MD, et al. Efficacy and safety of dabigatran compared with warfarin at different levels of international normalised ratio control for stroke prevention in atrial fibrillation: an analysis of the RE-LY trial. *Lancet.* 2010;376(9745):975-983.

79. Westreich D, Edwards JK. Invited commentary: every good randomization deserves observation. *Am J Epidemiol.* 2015;182(10):857-860.

80. van Onzenoort HA, Menger FE, Neef C, et al. Participation in a clinical trial enhances adherence and persistence to treatment: a retrospective cohort study. *Hypertension (Dallas, Tex : 1979).* 2011;58(4):573-578.

81. Granger CB, Lopes RD, Hanna M, et al. Clinical events after transitioning from apixaban versus warfarin to warfarin at the end of the Apixaban for Reduction in Stroke and Other Thromboembolic Events in Atrial Fibrillation (ARISTOTLE) trial. *American heart journal.* 2015;169(1):25-30.

82. Yao X, Abraham NS, Alexander GC, et al. Effect of adherence to oral anticoagulants on risk of stroke and major bleeding among patients with atrial fibrillation. *Journal of the American Heart Association.* 2016;5(2):e003074.

83. Lamberts M, Staerk L, Olesen JB, et al. Major Bleeding Complications and Persistence With Oral Anticoagulation in Non-Valvular Atrial Fibrillation: Contemporary Findings in Real-Life Danish Patients. *Journal of the American Heart Association.* 2017;6(2):e004517.

84. Jackevicius CA, Tsadok MA, Essebag V, et al. Early non-persistence with dabigatran and rivaroxaban in patients with atrial fibrillation. *Heart (British Cardiac Society).* 2017:heartjnl-2016-310672.

85. Johnson ME, Lefèvre C, Collings S-L, et al. Early real-world evidence of persistence on oral anticoagulants for stroke prevention in non-valvular atrial fibrillation: a cohort study in UK primary care. *BMJ Open.* 2016;6(9):e011471.

86. Miettinen OS. Standardization of risk ratios. *Am J Epidemiol.* 1972;96(6):383-388.

87. Cole SR, Stuart EA. Generalizing evidence from randomized clinical trials to target populations: The ACTG 320 trial. *Am J Epidemiol.* 2010;172(1):107-115.

88. Westreich D, Edwards JK, Lesko CR, Stuart E, Cole SR. Transportability of trial results using inverse odds of sampling weights. *American Journal of Epidemiology.* 2017.

89. Bareinboim E, Lee S, Honavar V, Pearl J. Transportability from multiple environments with limited experiments. Paper presented at: Advances in Neural Information Processing Systems2013.

90. Rothman KJ, Gallacher JE, Hatch EE. Why representativeness should be avoided. *International journal of epidemiology.* 2013;42(4):1012-1014.

91. Richiardi L, Pizzi C, Pearce N. Commentary: Representativeness is usually not necessary and often should be avoided. *International journal of epidemiology.* 2013;42(4):1018-1022.

92. Stang A, Jockel KH. Avoidance of representativeness in presence of effect modification. *International journal of epidemiology.* 2014;43(2):630-631.

93. Rothman K, Hatch E, Gallacher J. Representativeness is not helpful in studying heterogeneity of effects across subgroups. *International journal of epidemiology.* 2014;43(2):633-634.

94. Pressler TR, Kaizar EE. The use of propensity scores and observational data to estimate randomized controlled trial generalizability bias. *Statistics in Medicine.* 2013;32(20):3552-3568.

95. Stuart EA, Bradshaw CP, Leaf PJ. Assessing the generalizability of randomized trial results to target populations. *Prevention science : the official journal of the Society for Prevention Research.* 2015;16(3):475-485.

96. Swallow E, Kelley C, Ristovska L, et al. Daclatasvir +Asunaprevir Versus Sofosbuvir/Ledipasvir for The Treatment of Chronic Hepatitis C Genotype 1 In Japanese Patients: A Matching-Adjusted Indirect Comparison. *Value in Health.* 2015;18(7):A576.

97. Stuart EA, Bradshaw CP, Leaf PJ. Assessing the generalizability of randomized trial results to target populations. *Prevention Science.* 2015;16(3):475-485.

98. Katz DF, Maddox TM, Turakhia M, et al. Contemporary Trends in Oral Anticoagulant Prescription in Atrial Fibrillation Patients at Low to Moderate Risk of Stroke After Guideline-Recommended Change in Use of the CHADS2 to the CHA2DS2-VASc Score for Thromboembolic Risk Assessment: Analysis From the National Cardiovascular Data Registry's Outpatient Practice Innovation and Clinical Excellence Atrial Fibrillation Registry. *Circulation: Cardiovascular and Quality Outcomes.* 2017;10(5).

99. Biedermann JS, van Rein N, van den Besselaar AM, et al. Impact of point-of-care international normalized ratio monitoring on quality of treatment with vitamin K antagonists in non-self-monitoring patients: a cohort study. *Journal of Thrombosis and Haemostasis.* 2016;14(4):695-703.

100. Weitz JI, Semchuk W, Turpie AG, et al. Trends in Prescribing Oral Anticoagulants in Canada, 2008-2014. *Clinical therapeutics.* 2015;37(11):2506-2514.e2504.

101. Lang K, Bozkaya D, Patel AA, et al. Anticoagulant use for the prevention of stroke in patients with atrial fibrillation: findings from a multi-payer analysis. *BMC Health Serv Res.* 2014;14:329.

102. Patel PA, Zhao X, Fonarow GC, et al. Novel Oral Anticoagulant Use Among Patients with Atrial Fibrillation Hospitalized with Ischemic Stroke or Transient Ischemic Attack. *Circ Cardiovasc Qual Outcomes.* 2015;8(4):383-392.

103. Desai NR, Krumme AA, Schneeweiss S, et al. Patterns of Initiation of Oral Anticoagulants in Patients with Atrial Fibrillation— Quality and Cost Implications. *The American journal of medicine.*127(11):1075-1082.e1071.

104. Quarterly U.S. Sales Data for Pradaxa. https://www.drugs.com/stats/pradaxa. Accessed December 27, 2017.

105. Pearl J, Bareinboim E. External validity: From do-calculus to transportability across populations. *Statistical Science.* 2014;29(4):579-595.

106. Tamariz L, Harkins T, Nair V. A systematic review of validated methods for identifying venous thromboembolism using administrative and claims data. *Pharmacoepidemiology and Drug Safety.* 2012;21 Suppl 1:154-162.

107. Hernan MA, Robins JM. Per-Protocol Analyses of Pragmatic Trials. *New England Journal of Medicine.* 2017;377(14):1391-1398.

108. Andrade SE, Harrold LR, Tjia J, et al. A systematic review of validated methods for identifying cerebrovascular accident or transient ischemic attack using administrative data. *Pharmacoepidemiology and Drug Safety.* 2012;21 Suppl 1:100-128.

109. Wahl PM, Rodgers K, Schneeweiss S, et al. Validation of claims-based diagnostic and procedure codes for cardiovascular and gastrointestinal serious adverse events in a commercially-insured population. *Pharmacoepidemiology and Drug Safety.* 2010;19(6):596-603.

110. Cunningham A, Stein CM, Chung CP, Daugherty JR, Smalley WE, Ray WA. An Automated Database Case Definition for Serious Bleeding Related to Oral Anticoagulant Use. *Pharmacoepidemiology and Drug Safety.* 2011;20(6):560-566.

111. Lauffenburger JC, Balasubramanian A, Farley JF, et al. Completeness of prescription information in US commercial claims databases. *Pharmacoepidemiology and Drug Safety.* 2013;22(8):899-906.

112. Xing Y, Ma Q, Ma X, Wang C, Zhang D, Sun Y. CHADS(2) score has a better predictive value than CHA(2)DS(2)-VASc score in elderly patients with atrial fibrillation. *Clinical Interventions in Aging.* 2016;11:941-946.

113. Berry JD, Lloyd-Jones DM, Garside DB, Greenland P. Framingham risk score and prediction of coronary heart disease death in young men. *American heart journal.* 2007;154(1):80-86.

114. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine.* 2009;28(25):3083-3107.

115. Cole SR, Hernan MA. Adjusted survival curves with inverse probability weights. *Computer methods and programs in biomedicine.* 2004;75(1):45-49.

116. Li F, Morgan KL, Zaslavsky AM. Balancing covariates via propensity score weighting. *Journal of the American Statistical Association.* 2017:1-11.

117. Stuart EA, Cole SR, Bradshaw CP, Leaf PJ. The use of propensity scores to assess the generalizability of results from randomized trials. *Journal of the Royal Statistical Society Series A, (Statistics in Society).* 2011;174(2):369-386.

118. Faurot KR, Jonsson Funk M, Pate V, et al. Using claims data to predict dependency in activities of daily living as a proxy for frailty. *Pharmacoepidemiology and Drug Safety.* 2015;24(1):59-66.

119. Borgan Ø. Aalen–Johansen Estimator. *Encyclopedia of Biostatistics*: John Wiley & Sons, Ltd; 2005.

120. Tideman PA, Tirimacco R, St John A, Roberts GW. How to manage warfarin therapy. *Australian prescriber.* 2015;38(2):44-48.

121. Hernán MA. The hazards of hazard ratios. *Epidemiology (Cambridge, Mass).* 2010;21(1):13.

122. Poole C. On the origin of risk relativism. *Epidemiology.* 2010;21(1):3-9.

123. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology.* 1999;10(1):37-48.

124. Sturmer T, Rothman KJ, Glynn RJ. Insights into different results from different causal contrasts in the presence of effect-measure modification. *Pharmacoepidemiology and Drug Safety.* 2006;15(10):698-709.

125. Cain LE, Cole SR. Inverse probability-of-censoring weights for the correction of time-varying noncompliance in the effect of randomized highly active antiretroviral therapy on incident AIDS or death. *Statistics in Medicine.* 2009;28(12):1725-1738.

126. Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statistics in Medicine.* 2000;19(9):1141-1164.

127. Cuthbertson CC, Kucharska-Newton A, Faurot KR, et al. Controlling for Frailty in Pharmacoepidemiologic Studies of Older Adults: Validation of an Existing Medicare Claims-based Algorithm. *Epidemiology.* 2018;29(4):556-561.

128. Wong SL, Marshall LZ, Lawson KA. Direct oral anticoagulant prescription trends, switching patterns, and adherence in Texas Medicaid. *American Journal of Managed Care.* 2018;24(8 Spec No.):Sp309-sp314.

129. Shore S, Carey EP, Turakhia MP, et al. Adherence to dabigatran therapy and longitudinal patient outcomes: insights from the veterans health administration. *American heart journal.* 2014;167(6):810-817.

130. Patel MR, Hellkamp AS, Lokhnygina Y, et al. Outcomes of discontinuing rivaroxaban compared with warfarin in patients with nonvalvular atrial fibrillation: analysis from the ROCKET AF trial (Rivaroxaban Once-Daily, Oral, Direct Factor Xa Inhibition Compared With Vitamin K Antagonism for Prevention of Stroke and Embolism Trial in Atrial Fibrillation). *Journal of the American College of Cardiology.* 2013;61(6):651-658.

131. Lai CL, Chen HM, Liao MT, Lin TT. Dabigatran, Rivaroxaban, and Warfarin in the Oldest Adults with Atrial Fibrillation in Taiwan. *Journal of the American Geriatriatric Society.* 2018;66(8):1567-1574.

132. Simonsen L, Taylor RJ, Viboud C, Miller MA, Jackson LA. Mortality benefits of influenza vaccination in elderly people: an ongoing controversy. *The Lancet Infectious diseases.* 2007;7(10):658-666.

133. Mueller N, Murthy S, Tainter CR, et al. Can Sarcopenia Quantified by Ultrasound of the Rectus Femoris Muscle Predict Adverse Outcome of Surgical Intensive Care Unit Patients as well as Frailty? A Prospective, Observational Cohort Study. *Annals of Surgery.* 2016;264(6):1116-1124.

134. Cho J, Lee I, Park SH, et al. Socioeconomic Status, Frailty, and All-Cause Mortality in Korean Older Adults: A 3-Year Population-Based Prospective Study. *BioMed Research International.* 2017;2017:1903589-1903589.

135. Reynolds SL, Ghate SR, Sheer R, et al. Healthcare utilization and costs for patients initiating Dabigatran or Warfarin. *Health Quality of Life Outcomes.* 2017;15(1):128.

136. Waldo DR. Accuracy and Bias of Race/Ethnicity Codes in the Medicare Enrollment Database. *Health care financing review.* 2004;26(2):61-72.

137. Bhattacharya J, Vogt WB. Do instrumental variables belong in propensity scores? : National Bureau of Economic Research Cambridge, Mass., USA; 2007.

138. Toh S, Hernandez-Diaz S, Logan R, Robins JM, Hernan MA. Estimating absolute risks in the presence of nonadherence: an application to a follow-up study with baseline randomization. *Epidemiology.* 2010;21(4):528-539.

139. Hernan MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology.* 2008;19(6):766-779.

140. Lesko CR, Buchanan AL, Westreich D, Edwards JK, Hudgens MG, Cole SR. Generalizing Study Results: A Potential Outcomes Perspective. *Epidemiology.* 2017;28(4):553-561.

141. Westreich D, Edwards JK, Rogawski ET, Hudgens MG, Stuart EA, Cole SR. Causal Impact: Epidemiological Approaches for a Public Health of Consequence. *American Journal of Public Health.* 2016;106(6):1011.

142. Bengtson AM, Pence BW, Gaynes BN, et al. Improving Depression Among HIV-Infected Adults: Transporting the Effect of a Depression Treatment Intervention to Routine Care. *Journal of acquired immune deficiency syndromes (1999).* 2016;73(4):482-488.

143. Jameson JL, Longo DL. Precision medicine—personalized, problematic, and promising. *Obstetrical & gynecological survey.* 2015;70(10):612-614.

144. Hernán MA, VanderWeele TJ. Compound treatments and transportability of causal inference. *Epidemiology (Cambridge, Mass).* 2011;22(3):368.

145. Hong J-L, Jonsson Funk M, LoCasale R, et al. Generalizing Randomized Clinical Trial Results: Implementation and Challenges Related to Missing Data in the Target Population. *American Journal of Epidemiology.* 2017:kwx287-kwx287.

146. Susukida R, Crum RM, Ebnesajjad C, Stuart EA, Mojtabai R. Generalizability of findings from randomized controlled trials: application to the National Institute of Drug Abuse Clinical Trials Network. *Addiction (Abingdon, England).* 2017;112(7):1210-1219.

147. Webster-Clark MA, Sanoff HK, Stürmer T, Hinton SP, Lund JL. Diagnostic Assessment of Assumptions for External Validity: An Example Using Data in Metastatic Colorectal Cancer. *Epidemiology.* 2019;30(1):103-111.

148. Hernan MA, Hernandez-Diaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clinical trials (London, England).* 2012;9(1):48-55.

149. VanderWeele TJ. On the distinction between interaction and effect modification. *Epidemiology.* 2009;20(6):863-871.

150. King NB, Harper S, Young ME. Use of relative and absolute effect measures in reporting health inequalities: structured review. *Bmj.* 2012;345:e5774.

151. Sato T, Matsuyama Y. Marginal structural models as a tool for standardization. *Epidemiology.* 2003;14(6):680-686.

152. Lichtman JH, Leifheit-Limson EC, Goldstein LB. Centers for medicare and medicaid services medicare data and stroke research: goldmine or landmine? *Stroke.* 2015;46(2):598-604.

153. Lakshminarayan K, Larson JC, Virnig B, et al. Comparison of Medicare claims versus physician adjudication for identifying stroke outcomes in the Women's Health Initiative. *Stroke.* 2014;45(3):815-821.

154. Eapen ZJ, Hammill BG, Setoguchi S, et al. Who enrolls in the Medicare Part D prescription drug benefit program? Medication use among patients with heart failure. *Journal of the American Heart Association.* 2013;2(5):e000242.